201ab Quantitative methods L.04: Uncertainty, risk, and probability

Probability

- This section can get a bit mathy

- But, it is important:

  – All of statistics is based on probability

  – Probability is a useful tool for thinking about the world, because as far as we are concerned, the world is probabilistic, not deterministic.

Probability is the basis of statistics.

- Probability quantifies uncertainty (e.g., of inferences)
  used throughout data analysis

- Statistical model (+ parameters) + Probability
  = sampling/predicted (probability!) distribution
  used for prediction, NHST, etc.
  = probability of our data (likelihood) under parameters
  used for parameter estimation, model selection

Probability

- **Probability statements, philosophy**
- Basic probability rules
- Joint, conditional, and Bayes' rule.
- *Random variables and probability distributions*
  - Sampling to generate random values.
  - Probability mass (pmf) and density (pdf) functions
  - Cumulative probability density function (cdf)
  - Quantile function (inverse cdf)
- Expectation, Variance, and their rules
- Central limit theorem and the normal distribution.

Translating probability statements

- Probability $0.0 \le P \le 1.0$ ("proportion", "risk")

  Proportion used for observed counts, probability for predictions / unobserved events.
  "Expected/predicted" proportion: proportion you expect in a future count.

- Percent $0.0 \le \% \le 100.0$

  These are usually used to describe observed counts, but you could also use them to refer to chance.

  $\% = 100 * P$

- Odds $0 \le \textbf{odds} \le \infty$ (ratio; $10:1 = 10/1 = 10$)

  These are often used in gambling, and are helpful to describe how probability changes under different interventions (typically, odds are multiplied)

  $\text{odds} = P \,/\, (1-P)$

  $P = \text{odds} \,/\, (1+\text{odds})$

- Log-odds $-\infty \le \textbf{log-odds} \le \infty$

  We will deal with these a lot when considering logistic regression, because they behave linearly.

  $\text{log-odds} = \log(\text{odds}) = \log(P \,/\, (1-P))$

  $\text{odds} = \exp(\text{log-odds})$

  $P = \exp(\text{log-odds}) \,/\, (1+\exp(\text{log-odds}))$

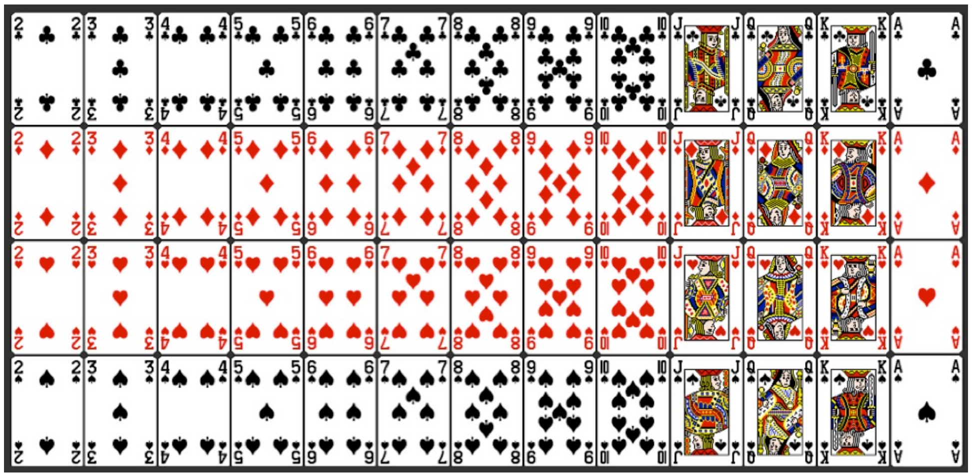
Pet peeve: Comparing probabilities

- Lifetime lung cancer “incidence”/“risk” for female...
  non-smoker: 0.003 (absolute probability) odds: 1:332
  smoker: 0.125 (absolute probability) odds: 1:7
- **Risk factor** for smoking: 0.125/0.003 = 41
  Risk is 41x higher for smokers.
  We could (probably shouldn’t) say the risk is 4000% greater.
  (relative risk -- proportional)
- Change in risk is 0.125-0.003 = 0.122
  Smoking yields a risk that is 0.122 units greater;
  Smoking increases risk by 12.2 percentage points
  (relative risk -- additive)
- Press often does not clarify which is being reported:
  Confusion between absolute and relative probability (and which is important).
  Confusion between additive and proportional probability differences
  Avoid this sort of confusion, figure out which one is being reported.

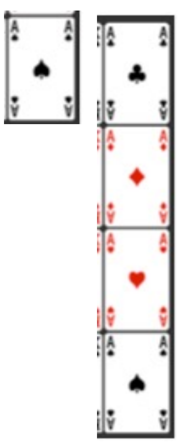
Probability

- Probability statements, philosophy
- **Basic probability rules**
- Joint, conditional, and Bayes' rule.
- *Random variables and probability distributions*
  - Sampling to generate random values.
  - Probability mass (pmf) and density (pdf) functions
  - Cumulative probability density function (cdf)
  - Quantile function (inverse cdf)
- Expectation, Variance, and their rules
- Central limit theorem and the normal distribution.

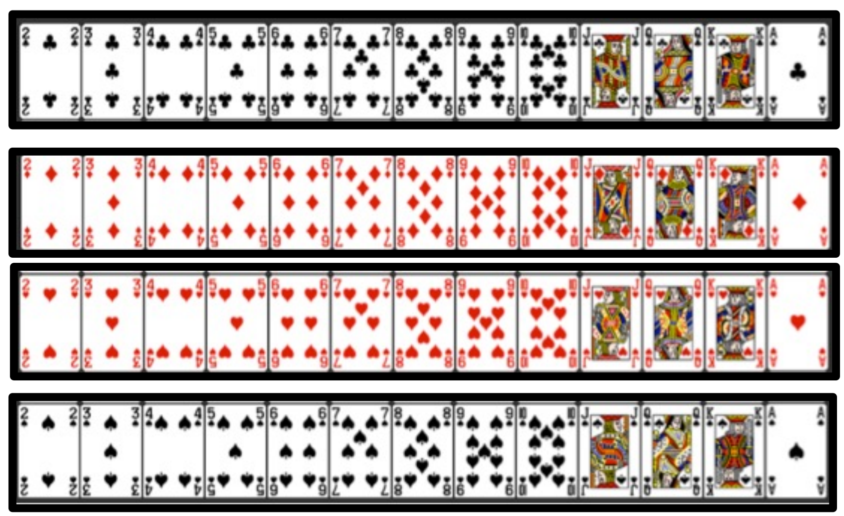
Technical jargon about "outcomes"



Sample/probability space
(the set of all possible outcomes)



"Events"
Groups of outcomes from the sample space



"Partition"
A set of "events" that are disjoint (non-overlapping) and cover the full sample space

Basic probability rules.

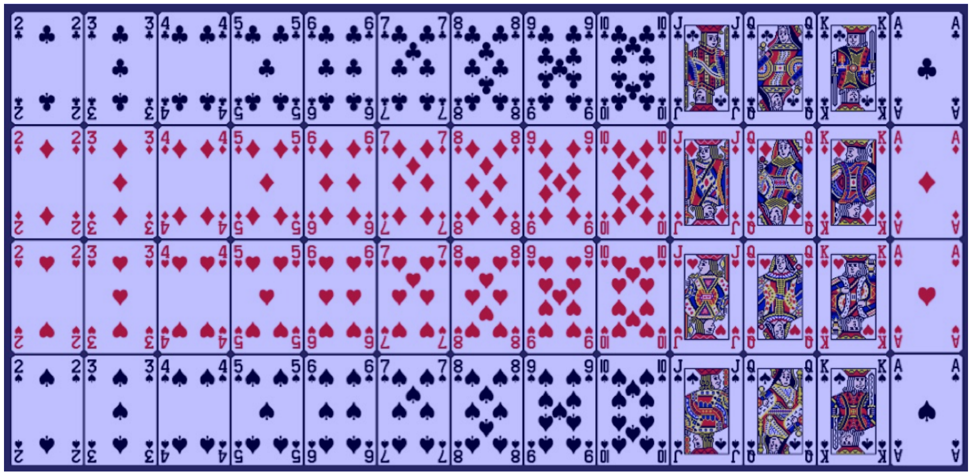
P(Ω) = 1

Ω: sample space

"P(anything)"

"you draw a card…"

Total area is 1



Basic probability rules.

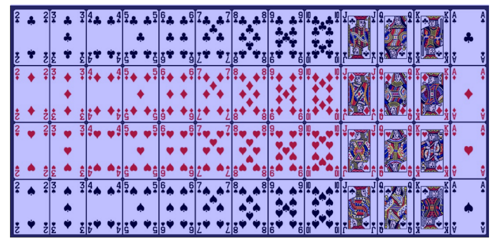
P(Ω) = 1

Ω: sample space

"P(anything)"

"you draw a card…"

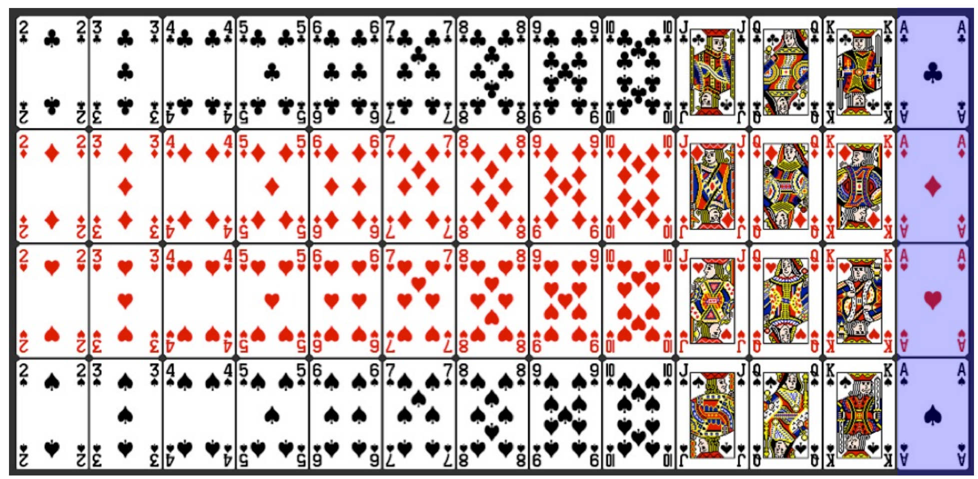
Total area is 1



0<=P(A)<=1

"Probability of A"

"you draw an ace…"



Basic probability rules.

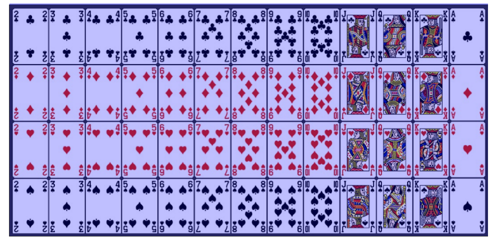
$P(\Omega) = 1$

Ω: sample space

"P(anything)"

"you draw a card…"

Total area is 1

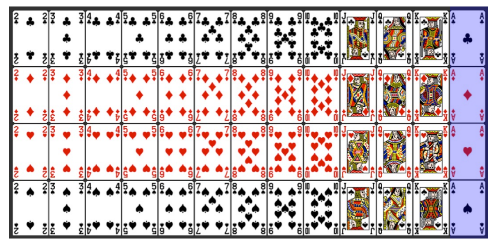


$0 <= P(A) <= 1$

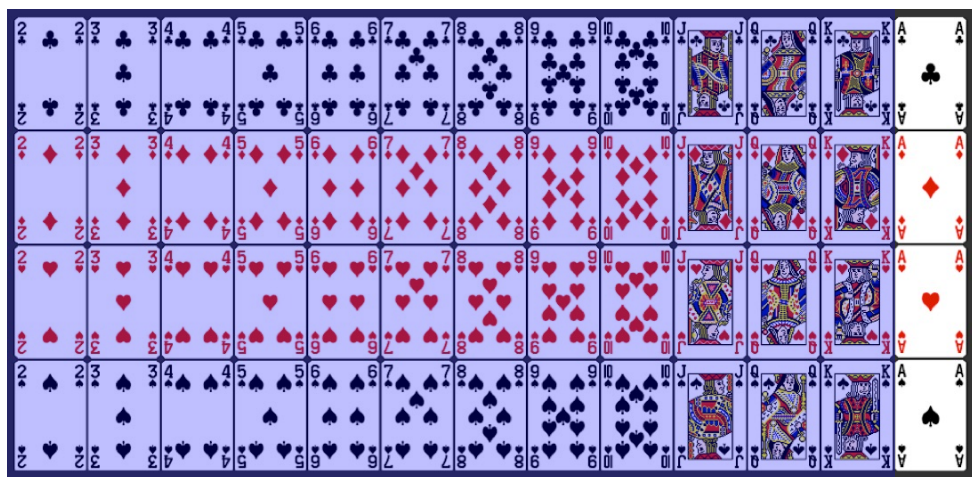
"Probability of A"

"you draw an ace…"

P(~A)



P(A)



$P(A)+P(\sim A) = 1$

$P(\sim A) = 1- P(A)$

"P(not A)"

"you do not draw an ace…"

Basic probability rules.

$P(\Omega) = 1$

Ω: sample space

"P(anything)"

$0 <= P(A) <= 1$

"Probability of A"

$P(A)+P(\sim A) = 1$

$P(\sim A) = 1 - P(A)$

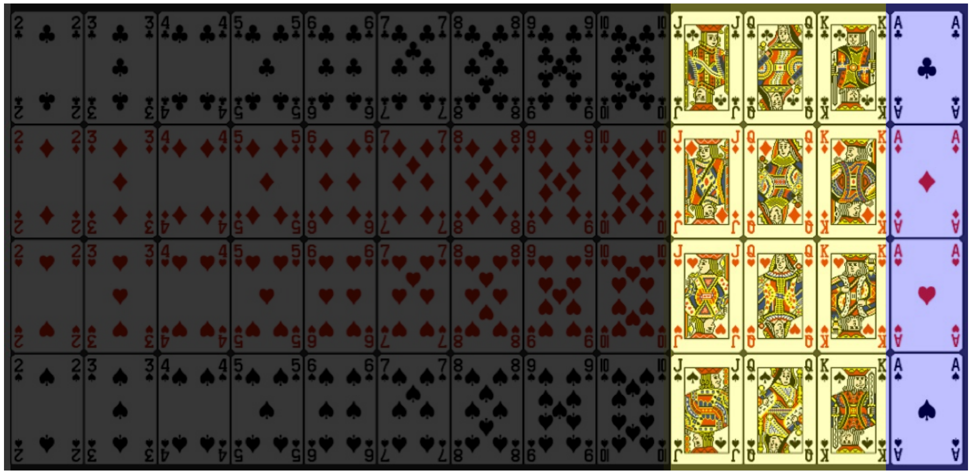
"P(not A)"

$P(A|B) = P(A\&B)/P(B)$

$P(A|B)$: conditional probability

"P(A given B)"

"you draw an ace, given that your card > 10..."



Basic probability rules.

$P(\Omega) = 1$

- Ω: sample space
- "P(anything)"

$0 <= P(A) <= 1$

- "Probability of A"

$P(A)+P(\sim A) = 1$

- $P(\sim A) = 1 - P(A)$
- "P(not A)"

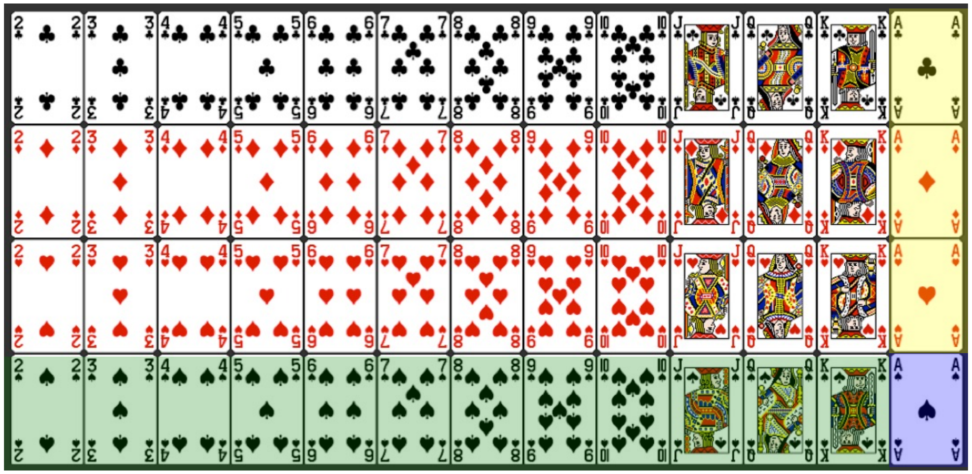
$P(A|B) = P(A\&B)/P(B)$

- P(A|B): conditional probability
- "P(A given B)"

$P(A\&B) = P(B|A)*P(A)$

- P(A&B): joint probability conjunction
- "P(A and B)"

"you draw the ace of spades…"



Basic probability rules.

$P(\Omega) = 1$

Ω: sample space

"P(anything)"

$0 <= P(A) <= 1$

"Probability of A"

$P(A)+P(\sim A) = 1$

$P(\sim A) = 1 - P(A)$

"P(not A)"

$P(A|B) = P(A\&B)/P(B)$

$P(A|B)$: conditional probability

"P(A given B)"

$P(A\&B) = P(B|A)*P(A)$

$P(A\&B)$: joint probability

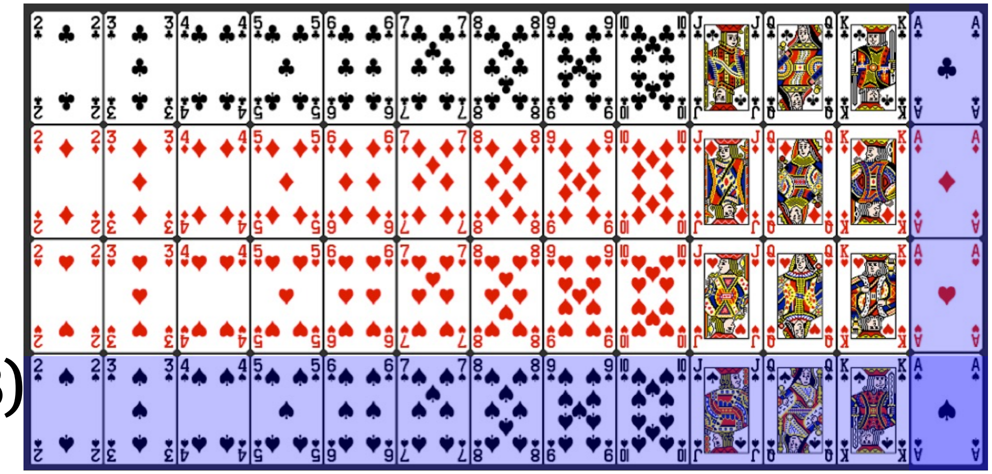
conjunction

"P(A and B)"

$P(A \vee B) = P(A) + P(B) - P(A\&B)$

disjunction

"P(A or B)"



"you draw an ace or a spade…"

Basic probability rules.

$P(\Omega) = 1$

Ω: sample space

"P(anything)"

$0 <= P(A) <= 1$

"Probability of A"

$P(A)+P(\sim A) = 1$

$P(\sim A) = 1 - P(A)$

"P(not A)"

$P(A|B) = P(A\&B)/P(B)$

$P(A|B)$: conditional probability

"P(A given B)"

$P(A\&B) = P(B|A)*P(A)$

$P(A\&B)$: joint probability

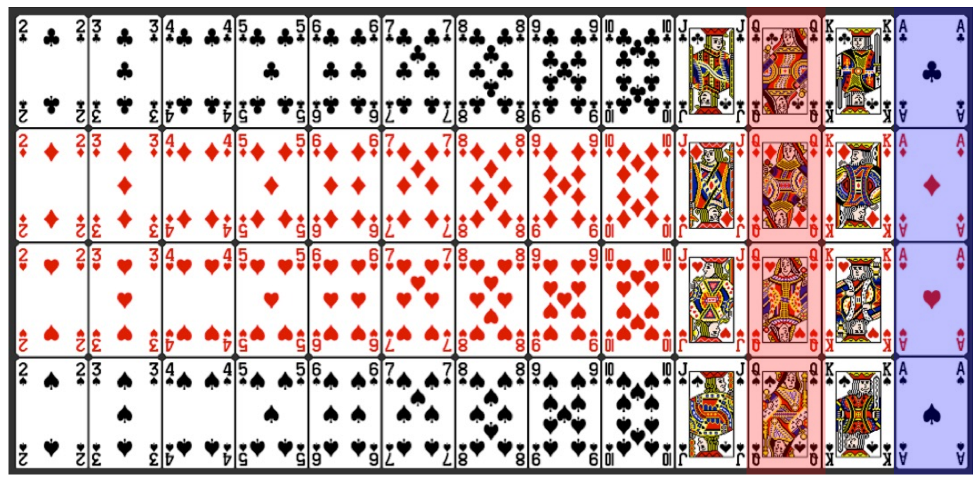
conjunction

"P(A and B)"

$P(A \vee B) = P(A) + P(B) - P(A\&B)$

disjunction

"P(A or B)"



"ace", "queen"

Disjoint events

iff $P(A\&B)=0$

thus $P(A \vee B) = P(A) + P(B)$

Basic probability rules.

$P(\Omega) = 1$

Ω: sample space

"P(anything)"

$0 <= P(A) <= 1$

"Probability of A"

$P(A)+P(\sim A) = 1$

$P(\sim A) = 1 - P(A)$

"P(not A)"

$P(A|B) = P(A\&B)/P(B)$

$P(A|B)$: conditional probability

"P(A given B)"

$P(A\&B) = P(B|A)*P(A)$

$P(A\&B)$: joint probability

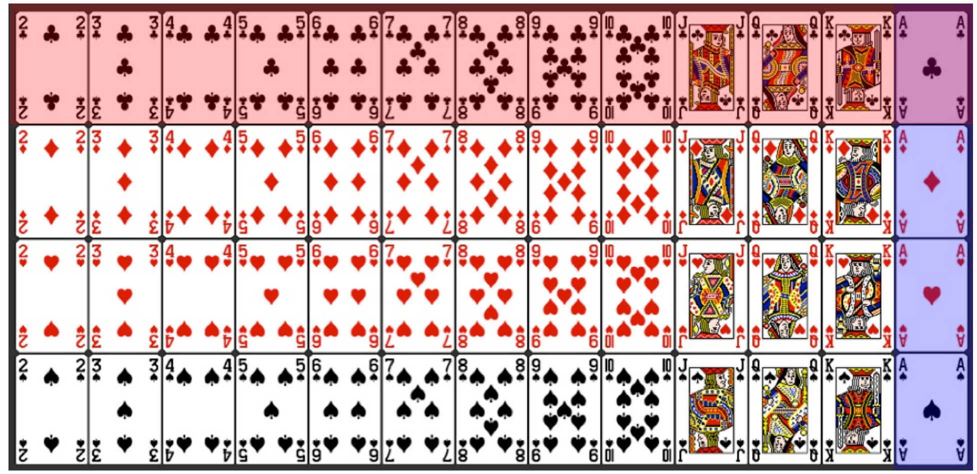
conjunction

"P(A and B)"

$P(A \vee B) = P(A) + P(B) - P(A\&B)$

disjunction

"P(A or B)"



"ace", "clubs"

Independent events

iff $P(A\&B)=P(A)*P(B)$

thus $P(A|B)=P(A)$

Disjoint events

iff $P(A\&B)=0$

thus $P(A \vee B) = P(A) + P(B)$

Basic probability rules.

$P(\Omega) = 1$

Ω: sample space

"P(anything)"

$0 <= P(A) <= 1$

"Probability of A"

$P(A)+P(\sim A) = 1$

$P(\sim A) = 1 - P(A)$

"P(not A)"

$P(A|B) = P(A\&B)/P(B)$

P(A|B): conditional probability

"P(A given B)"

$P(A\&B) = P(B|A)*P(A)$

P(A&B): joint probability

conjunction

"P(A and B)"

$P(A \vee B) = P(A) + P(B) - P(A\&B)$

disjunction

"P(A or B)"

Bayes Rule

$P(A|B)=P(B|A)\,P(A) / P(B)$

follows from def. of conditional prob.

Law of total probability

$P(B) = \Sigma_n [P(B \,\&\, A_n)]$

P(B): marginal probability

(given that A_ns are a partition)

Independent events

iff $P(A\&B)=P(A)*P(B)$

thus $P(A|B)=P(A)$

Disjoint events

iff $P(A\&B)=0$

thus $P(A \vee B) = P(A) + P(B)$

Law of total probability

P(ace) = P(ace & spade) +
P(ace & heart) +
P(ace & club) +
P(ace & diamond)

Law of total probability

$P(B) = \Sigma_n [P(B \& A_n)]$

P(B): marginal probability

(given that A_ns are a partition)

A needs to be a "partition" meaning the set of A_ns
(a) are mutually exclusive, and
(b) cover the whole sample space (sum($P(A_n)$)=1)

An example that doesn't work because it is not a partition:
P(ace) **not=** P(ace & diamond) + P(ace & not a face card) + P(ace & bicycle deck)

Basic probability rules.

$P(\Omega) = 1$

Ω: sample space

"P(anything)"

$0 <= P(A) <= 1$

"Probability of A"

$P(A)+P(\sim A) = 1$

$P(\sim A) = 1 - P(A)$

"P(not A)"

$P(A|B) = P(A\&B)/P(B)$

$P(A|B)$: conditional probability

"P(A given B)"

$P(A\&B) = P(B|A)*P(A)$

$P(A\&B)$: joint probability

conjunction

"P(A and B)"

$P(A \vee B) = P(A) + P(B) - P(A\&B)$

disjunction

"P(A or B)"

Bayes Rule

$P(A|B)=P(B|A)\,P(A) / P(B)$

follows from def. of conditional prob.

Law of total probability

$P(B) = \Sigma_n [\, P(B \,\&\, A_n) \,]$

P(B): marginal probability

(given that A_ns are a partition)

Independent events

iff $P(A\&B)=P(A)*P(B)$

thus $P(A|B)=P(A)$

Disjoint events

iff $P(A\&B)=0$

thus $P(A \vee B) = P(A) + P(B)$

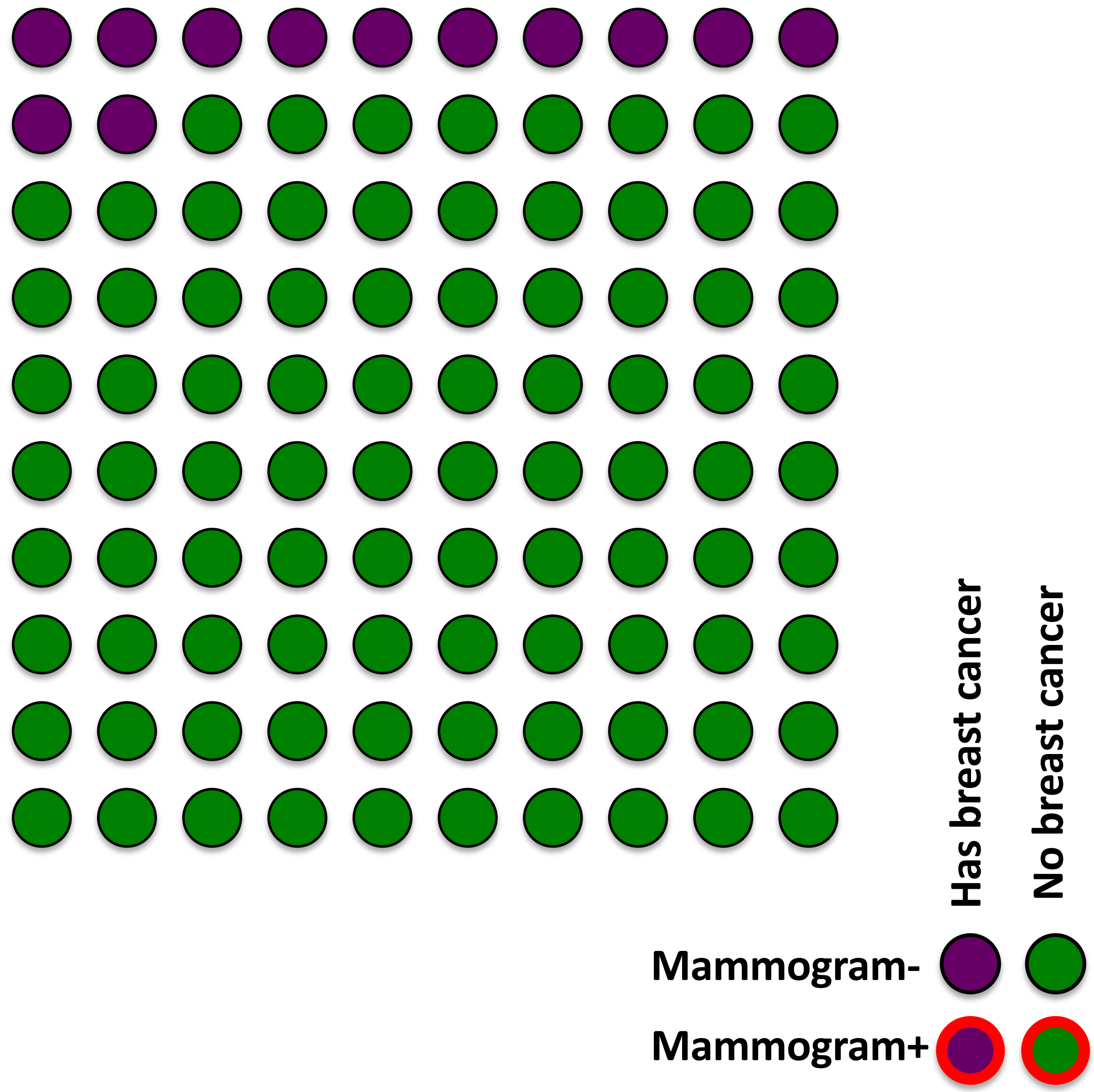
Probability

- Probability statements, philosophy
- Basic probability rules
- **Joint, conditional, and Bayes' rule.**
- *Random variables and probability distributions*
  - Sampling to generate random values.
  - Probability mass (pmf) and density (pdf) functions
  - Cumulative probability density function (cdf)
  - Quantile function (inverse cdf)
- Expectation, Variance, and their rules
- Central limit theorem and the normal distribution.

Mammogram contingency

Suppose:

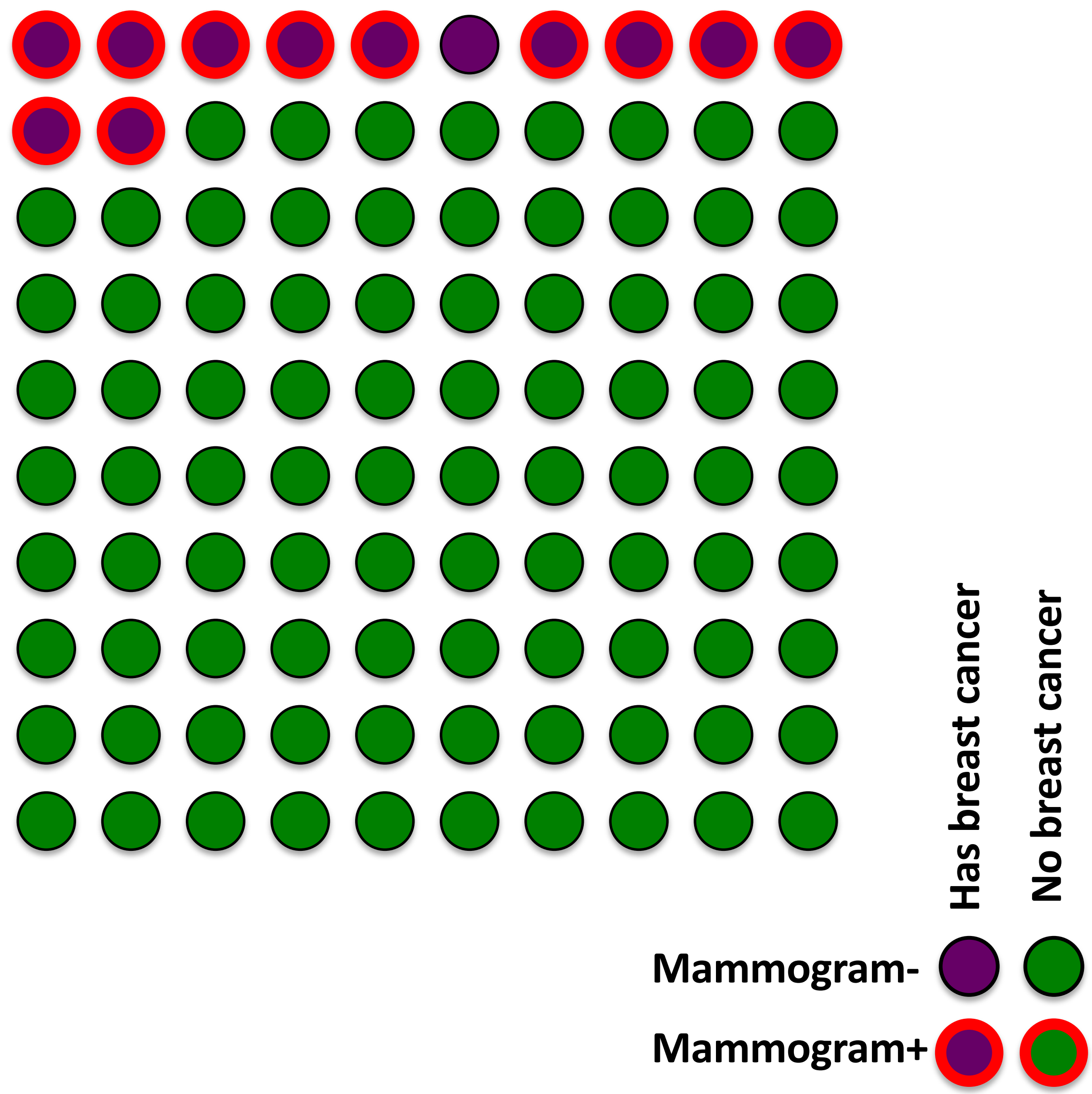
- **12/100 women have breast cancer**



Mammogram contingency

Suppose:

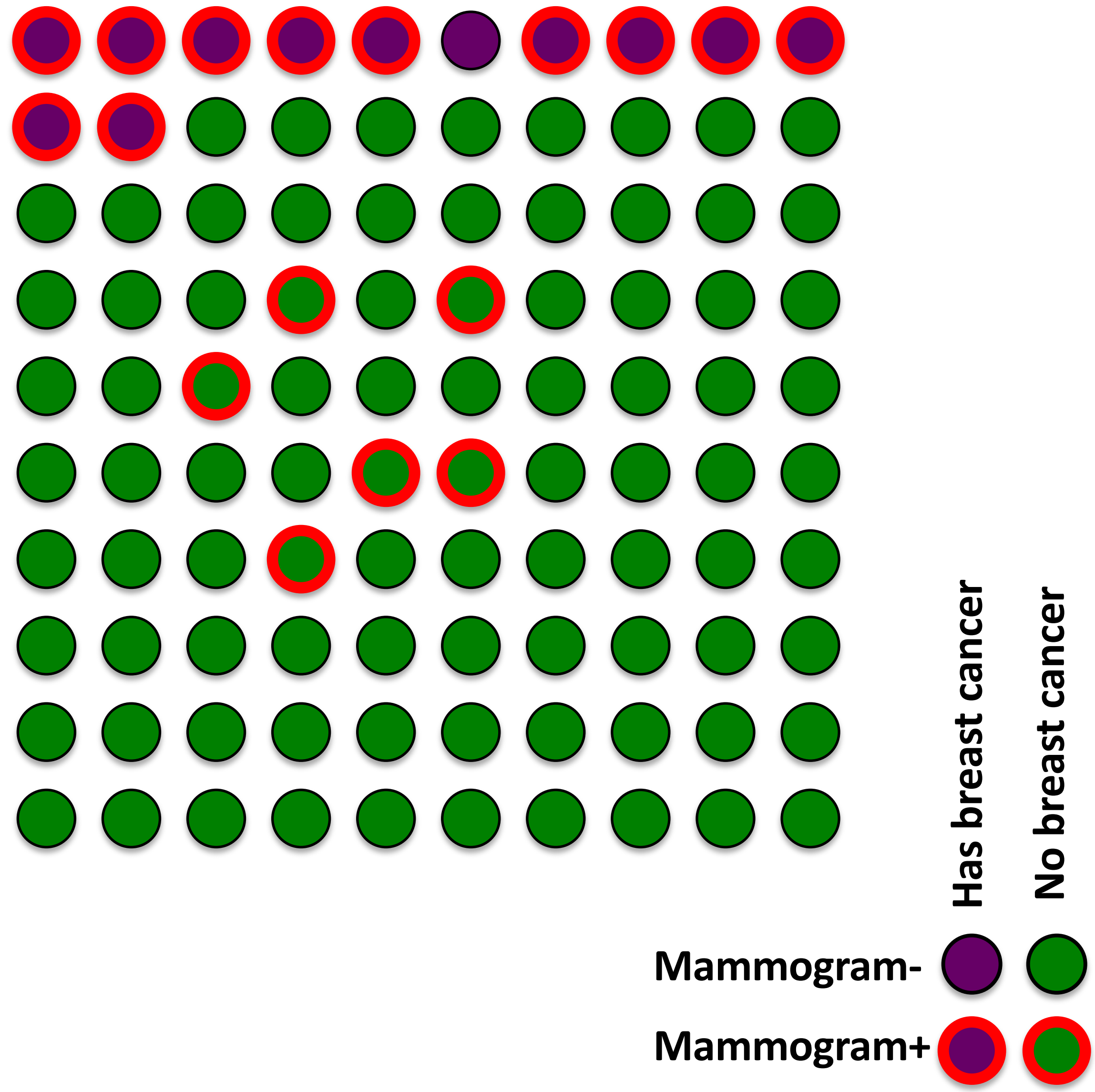
- 12/100 women have breast cancer
- A mammogram will detect breast cancer 90% of the time.



Mammogram contingency

Suppose:

- **12/100 women have breast cancer**
- **A mammogram will detect breast cancer 90% of the time.**
- **A mammogram will *falsely detect* breast cancer 7% of the time.**

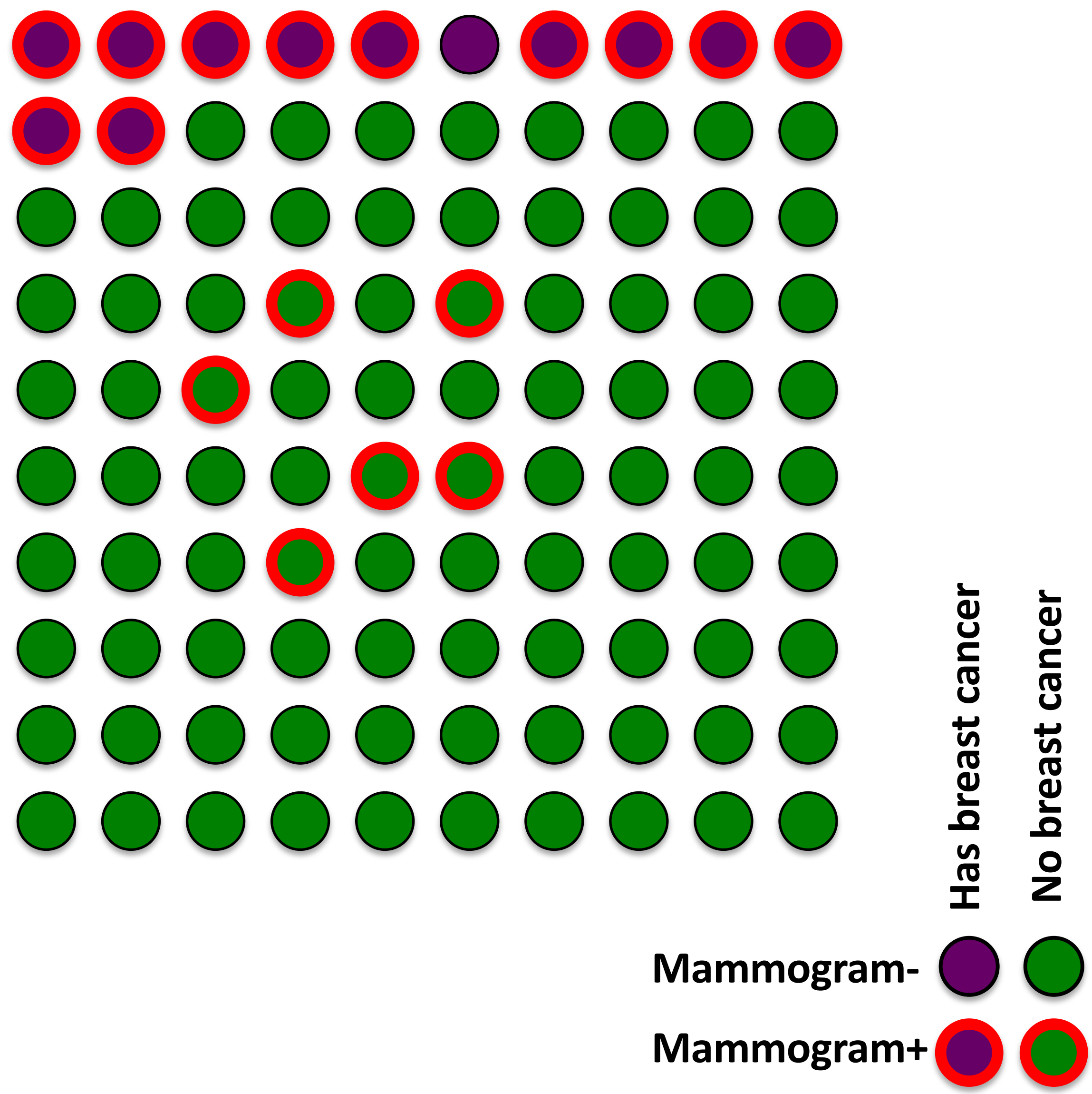


Mammogram contingency

Suppose:

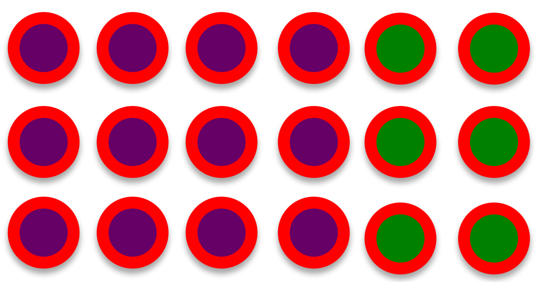
- 12/100 women have breast cancer
- A mammogram will detect breast cancer 90% of the time.
- A mammogram will *falsely detect* breast cancer 7% of the time.

What proportion of women with a positive mammogram have breast cancer?

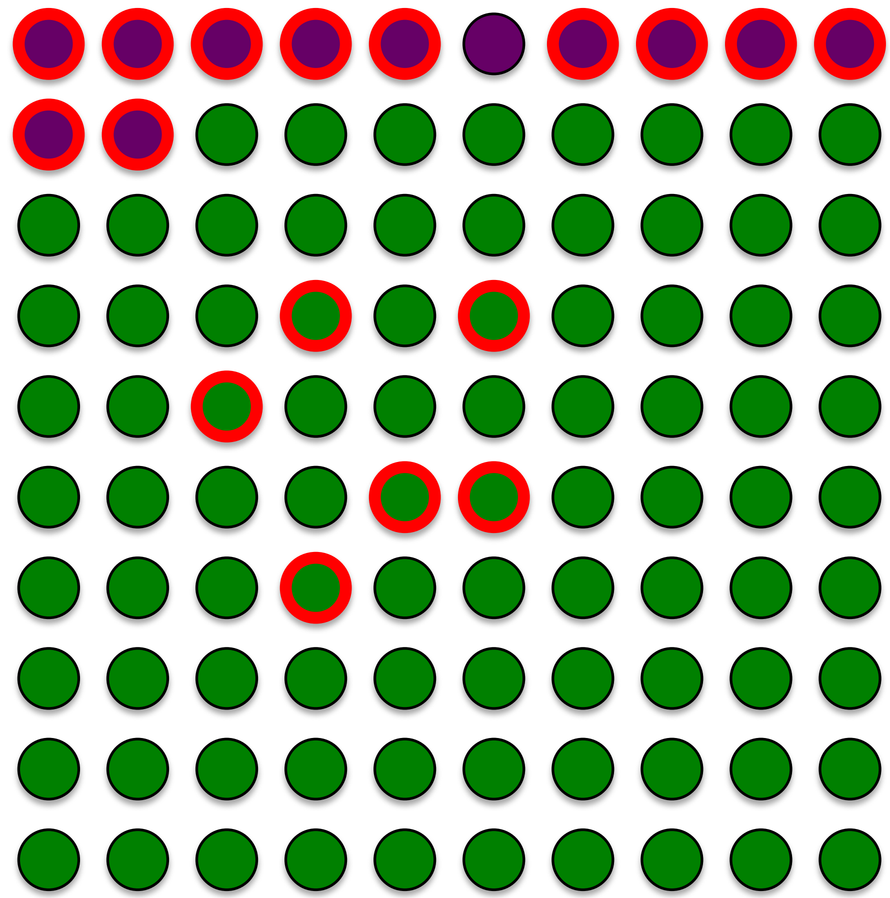


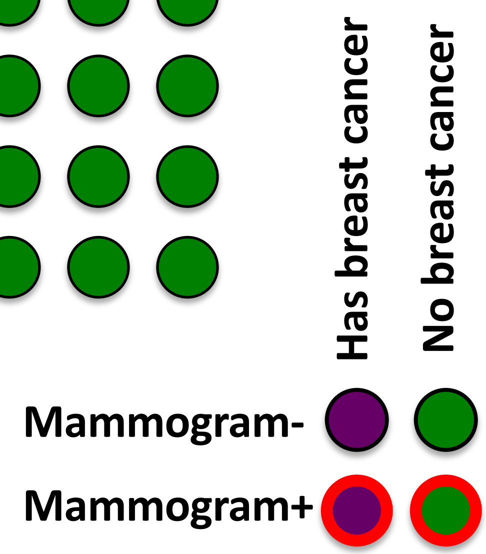
Mammogram contingency

What proportion of women with a positive mammogram have breast cancer?



~ 12/18 women with a positive mammogram have breast cancer.





Mammogram contingency

Suppose:

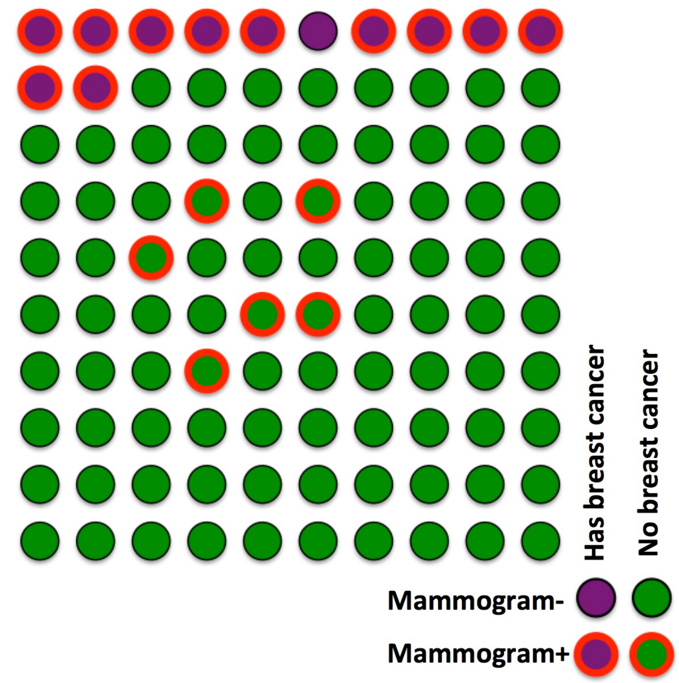
- 12/100 women have breast cancer

  P(cancer) = 0.12

  ***implied:* P(no cancer) = 0.88**
- A mammogram will detect breast cancer 90% of the time.

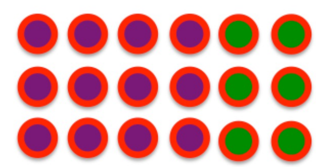
  P(mammogram+ | cancer) = 0.9
- A mammogram will *falsely detect* breast cancer 7% of the time.

  P(mammogram+ | no cancer) = 0.07



What proportion of women with a positive mammogram have breast cancer?

P(cancer | mammogram+) = ?



This is a Bayes rule question: going from one set of conditional probabilities, P(m+ | cancer), to their inverse: P(cancer | m+).

Bayes rule inverts conditionals

$$P(A|B) = P(B|A)\ P(A)\ /\ P(B)$$

Usually using the law of total probability to obtain P(B)

$P(B) = \text{sum}_A\ [\ P(B|A)P(A)\ ]$

Mammogram contingency

Marginal probability:
P(Cancer)

$$P(A|B) = \frac{P(B|A)\,\boxed{P(A)}}{P(B)}$$

	Mammogram positive	Mammogram negative	
No Cancer			9900 *0.99*
Cancer			100 *0.01*
			10,000 *1.00*

Mammogram contingency

Conditional probability:
P(Mammorgram | no cancer)

$$P(A|B) = \frac{\boxed{P(B|A)}\,P(A)}{P(B)}$$

	Mammogram positive	Mammogram negative	
No Cancer	*0.07*	*0.93*	9900 *0.99*
Cancer			100 *0.01*
			10,000 *1.00*

Mammogram contingency

Joint probability:
P(Mammogram & no cancer)

$\boxed{P(B \,\&\, A)}$

$$P(A|B) = \frac{\boxed{P(B|A)\, P(A)}}{P(B)}$$

	Mammogram positive	Mammogram negative	
No Cancer	693 *0.069* *0.07*	9207 *0.921* *0.93*	9900 *0.99*
Cancer			100 *0.01*
			10,000 *1.00*

Mammogram contingency

$$P(A|B) = \frac{\boxed{P(B|A)}\,P(A)}{P(B)}$$

Conditional probability:
P(Mammorgram | cancer)

	Mammogram positive	Mammogram negative	
No Cancer	693 *0.069* *0.07*	9207 *0.921* *0.93*	9900 *0.99*
Cancer	*0.8*	*0.2*	100 *0.01*
			10,000 *1.00*

Mammogram contingency

Joint probability:
P(Mammogram & cancer)

$\boxed{P(B \,\&\, A)}$

$$P(A|B) = \frac{\boxed{P(B|A)\ P(A)}}{P(B)}$$

	Mammogram positive	Mammogram negative	
No Cancer	693 *0.069* *0.07*	9207 *0.921* *0.93*	9900 *0.99*
Cancer	80 *0.008* *0.8*	20 *0.002* *0.2*	100 *0.01*
			10,000 *1.00*

Mammogram contingency

$$P(B) = \Sigma_A \, P(B \,\&\, A)$$

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Marginal probability: P(Mammogram)

	Mammogram positive	Mammogram negative	
No Cancer	693 *0.069* *0.07*	9207 *0.921* *0.93*	9900 *0.99*
Cancer	80 *0.008* *0.8*	20 *0.002* *0.2*	100 *0.01*
	773 *0.077*	9227 *0.923*	10,000 *1.00*

Mammogram contingency

$$\boxed{P(A|B)} = \frac{P(B|A)\,P(A)}{P(B)}$$

Conditional probability:
P(cancer | Mamm. negative)

	Mammogram positive	Mammogram negative	
No Cancer	693 *0.069* *0.07*	9207 *0.921* *0.93* *0.998*	9900 *0.99*
Cancer	80 *0.008* *0.8*	20 *0.002* *0.2* *0.002*	100 *0.01*
	773 *0.077*	9227 *0.923*	10,000 *1.00*

Mammogram contingency

$$\boxed{P(A|B)} = \frac{P(B|A)\,P(A)}{P(B)}$$

Conditional probability:
P(cancer | Mamm. positive)

	Mammogram positive	Mammogram negative	
No Cancer	693 *0.069* *0.07* *0.897*	9207 *0.921* *0.93* *0.998*	9900 *0.99*
Cancer	80 *0.008* *0.8* *0.103*	20 *0.002* *0.2* *0.002*	100 *0.01*
	773 *0.077*	9227 *0.923*	10,000 *1.00*

Mammogram contingency

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

	Mammogram positive	Mammogram negative	
No Cancer	693 *0.069* *0.07* *0.897*	9207 *0.921* *0.93* *0.998*	9900 *0.99*
Cancer	80 *0.008* *0.8* *0.103*	20 *0.002* *0.2* *0.002*	100 *0.01*
	773 *0.077*	9227 *0.923*	10,000 *1.00*

Make the full contingency table

- P(PCR+ | covid) = 0.91
  (sensitivity, recall, hit-rate, true positive rate)
- P(PCR+ | ~covid) = 0.023
  (false positive rate)
- P(covid) = 14000/1410000 ?
  (~1000 daily cases * ~14 day active period?)
- What's P(covid | PCR+)?

https://virologyj.biomedcentral.com/articles/10.1186/s12985-021-01489-0

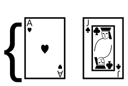
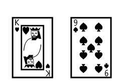
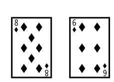
https://www.icd10monitor.com/false-positives-in-pcr-tests-for-covid-19

(15 min of googling, I would welcome more accurate numbers!)

Probability

- Probability statements, philosophy
- Basic probability rules
- Joint, conditional, and Bayes' rule.
- *Random variables and probability distributions*
  - Sampling to generate random values.
  - Probability mass (pmf) and density (pdf) functions
  - Cumulative probability density function (cdf)
  - Quantile function (inverse cdf)
- Expectation, Variance, and their rules
- Central limit theorem and the normal distribution.

Random variables

- Assign a value to each possible outcome.
  The value we get is a **"random variable"**
  - Random phenomenon ("experiment"): flip a coin.
    Outcomes: {Head, Tail}
    R.V. $X = \{1 \text{ if Head}; 0 \text{ if Tail}\}$
  - Random phenomenon: roll two dice.
    Outcomes: {(1,1); (1,2); (1,3); ... (6,5); (6,6)}
    R.V. A = die.1 + die.2
    R.V. B = die.1 * die.2
    R.V. C = die.1 – die.2
    R.V. D = {1 if die.1+die.2 is prime; 0 otherwise}
  - Random phenomenon: draw two cards from a deck.
    Outcomes: {, , , ...}
    R.V. best blackjack sum {4, ..., 21}

Random variables

- Assign a value to each possible outcome.
  The value we get is a **“random variable”**

- The possible values of the random variable are it’s *support.*
  - flip a coin 10 times
    X = # of heads (random variable)
    Support: {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10}
    no other values of X are possible

Random variables

- Assign a value to each possible outcome.
  The value we get is a **“random variable”**

- The possible values of the random variable are it’s *support.*
- We can characterize the random variable a few ways:
  - **Sampling process**
  - **Probability distribution function**
  - **Cumulative distribution function**
  - **Quantile function**
  - **Moments** (moment generating function)

Probability

- Probability statements, philosophy
- Basic probability rules
- Joint, conditional, and Bayes' rule.
- *Random variables and probability distributions*
  - **Sampling to generate random values.**
  - Probability mass (pmf) and density (pdf) functions
  - Cumulative probability density function (cdf)
  - Quantile function (inverse cdf)
- Expectation, Variance, and their rules
- Central limit theorem and the normal distribution.

Binomial random variable: sampling

- We flip a fair coin 3 times.
  Outcomes: {HHH, HHT, HTH, THH, …}
  X = number of times we got heads. {0,1,2,3}

- How do we sample this random variable?
  - Sampling: stochastically generate values with frequency proportional to their probability under this RV.
  - Sampling can be accomplished by stochastically simulating the experiment and evaluating the RV, or various procedures/algorithms that have the same properties, but are more efficient.

Binomial random variable: sampling

- We flip a fair coin 3 times.
  X = number of times we got heads. {0,1,2,3}
- How do we sample this random variable?

```
> flip()
[1] "T"
> flip()
[1] "H"
> flip()
[1] "T"
> flip()
[1] "H"
```

```
flip = function(){
  ifelse(runif(1)<0.5, 'H', 'T')
}
```

```
flip.n = function(n){
  replicate(n, flip())
}
```

```
> flip.n(3)
[1] "T" "T" "H"
> flip.n(3)
[1] "H" "H" "H"
> flip.n(3)
[1] "T" "H" "T"
> flip.n(3)
[1] "H" "H" "T"
```

```
evaluate.X = function(outcome){
  sum(outcome == 'H')
}
```

```
> evaluate.X(c('H', 'H', 'T'))
[1] 2
```

Binomial random variable: sampling

- We flip a fair coin 3 times.
  X = number of times we got heads. {0,1,2,3}
- How do we sample this random variable?

```
flip = function(){
  ifelse(runif(1)<0.5, 'H', 'T')
}
```

```
flip.n = function(n){
  replicate(n, flip())
}
```

```
evaluate.X = function(outcome){
  sum(outcome == 'H')
}
```

```
> evaluate.X(flip.n(3))
[1] 1
> evaluate.X(flip.n(3))
[1] 2
> evaluate.X(flip.n(3))
[1] 0
```

Binomial random variable: sampling

- We flip a fair coin 3 times.
  X = number of times we got heads. {0,1,2,3}
- How do we sample this random variable?

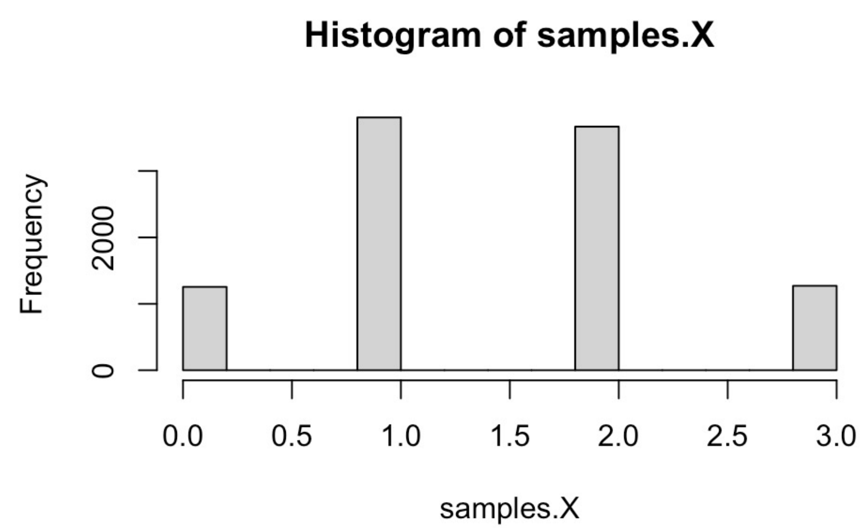
```
flip = function(){
  ifelse(runif(1)<0.5, 'H', 'T')
}
```

```
flip.n = function(n){
  replicate(n, flip())
}
```

```
evaluate.X = function(outcome){
  sum(outcome == 'H')
}
```

```
> samples.X = replicate(10000, evaluate.X(flip.n(3)))
> samples.X
  [1] 1 0 1 1 1 1 0 0 2 3 2 2 1 1 0 1 1 0 1 2 3 3
 [23] 2 0 1 2 1 0 1 3 2 1 0 3 1 2 2 2 0 1 1 2 0 2
 [45] 1 1 3 0 1 3 2 2 1 1 1 1 1 2 1 1 3 2 2 1 0 3
```

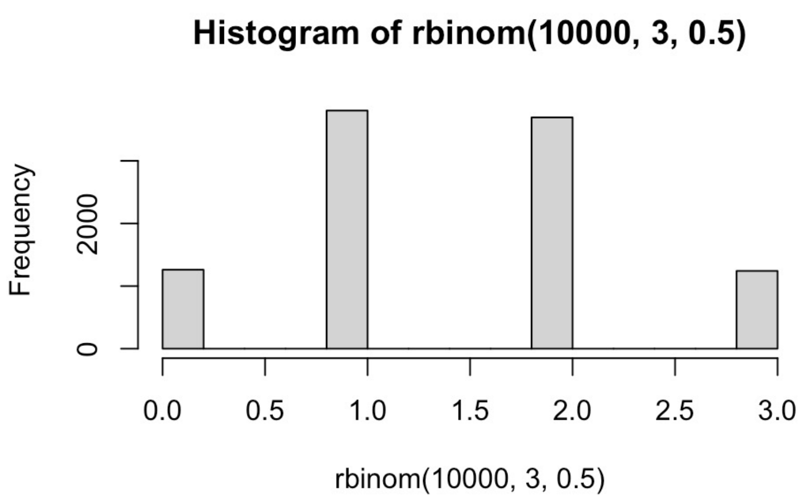
```
> hist(samples.X)
```



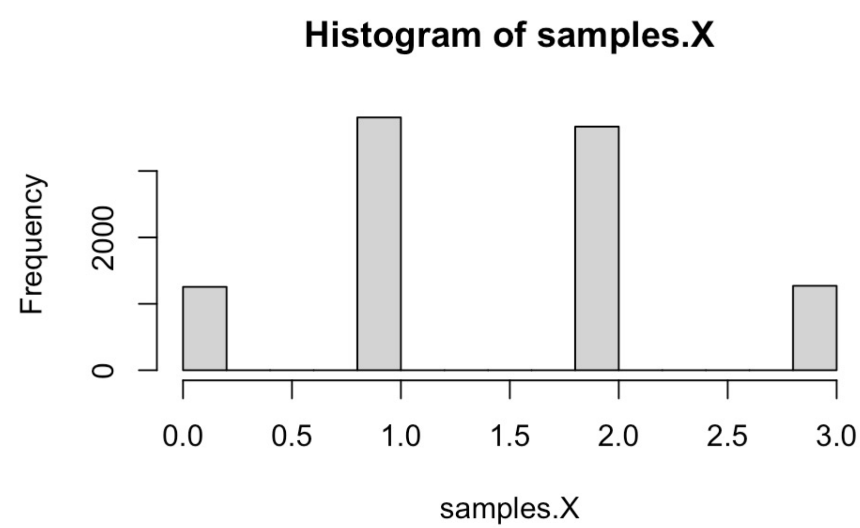
Binomial random variable: sampling

- We flip a fair coin 3 times.
  X = number of times we got heads. {0,1,2,3}
- How do we sample this random variable?

```
> hist(rbinom(10000, 3, 0.5))
```




```
> hist(samples.X)
```



Binomial random variable: sampling

- We flip a fair coin 3 times.
  X = number of times we got heads. {0,1,2,3}
- How do we sample this random variable?
  - This is a special kind of random variable for which we have a well-defined, named, probability distribution:
  - This is a *Binomial* random variable with *parameters*:
    n = 3 *"size"* (we flipped a fair coin 3 times)
    p = 0.5 *"prob"* (probability of heads was fair; i.e., 0.5)
  - So we can use the built-in R functions to sample it:
    rbinom(*number_of_samples*, *size, prob*)

rbinom(n, size, p)

n: number of random draws/samples

size: number of attempts (e.g., number of coin flips)

p: probability of success (e.g., prob coin lands heads)

Returns: a vector of xs (length n), each one representing a random draw of a binomial variable (e.g., number of heads out of *size* flips, each coming up heads with probability *p*.)

Sampling a random variable

- In R: **r***

```
rbinom(n, …)
rbeta(n, …)
rnorm(n, …)
rgeom(n, …)
rf(n, …)
rt(n, …)
```

- Each returns *n* randomly sampled values.

- Each distribution has its own parameters (like size and prob for Binomial).

Sampling to solve prob. problems

- If *x* is a vector of *n* samples from the random variable *X*, then (if *n* is large) we can…
  - Approximate *Pr(X=?)* as the frequency with which *x=?*
    In R notation: `sum(x==?)/n` or `mean(x==?)`
    [Borel's law of large numbers]
  - Approximate expectations (which we haven't yet covered) of *f(X)* as the average of *f(x)*
    In R notation: `sum(f(x))/n` or `mean(f(x))`
    [Monte Carlo theorem]
- So, we can estimate the frequency of getting 3 heads out of 10 from a fair coin as (where n is something big):

```
mean( rbinom(n,10,0.5)==3 )
```

Probability

- Probability statements, philosophy
- Basic probability rules
- Joint, conditional, and Bayes' rule.
- *Random variables and probability distributions*
  - Sampling to generate random values.
  - **Probability mass (pmf) and density (pdf) functions**
  - Cumulative probability density function (cdf)
  - Quantile function (inverse cdf)
- Expectation, Variance, and their rules
- Central limit theorem and the normal distribution.

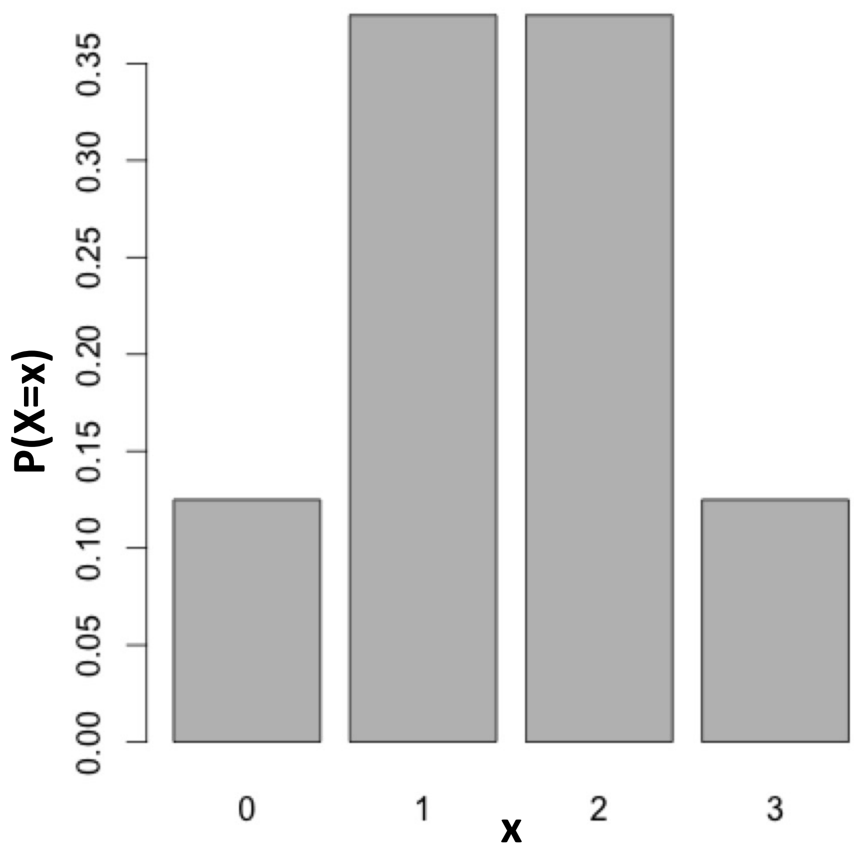
Binomial random variable: probability dist.

- We flip a fair coin 3 times.
  X = number of times we got heads. {0,1,2,3}
- What is the probability distribution of X?
  - The probability distribution describes the probability of seeing any particular value of X.
  - Here the probability distribution would specify the probabilities P(X=0), P(X=1), P(X=2), P(X=3)
  - In general,
    P(X=x) = sum(P(outcome))
    for all outcomes where X=x

Binomial random variable: probability dist.

- We flip a fair coin 3 times.
  X = number of times we got heads. {0,1,2,3}
- What is the probability distribution of X?
  P(X=0) = P(TTT) = 1*0.5^3 = 0.125
  P(X=1) = P(TTH) + P(THT) + P(HTT) = 3*0.5^3 = 0.375
  P(X=2) = P(THH) + P(HHT) + P(HTH) = 3*0.5^3 = 0.375
  P(X=3) = P(HHH) = 1*0.5^3 = 0.125

Binomial random variables



Binomial random variable: probability dist.

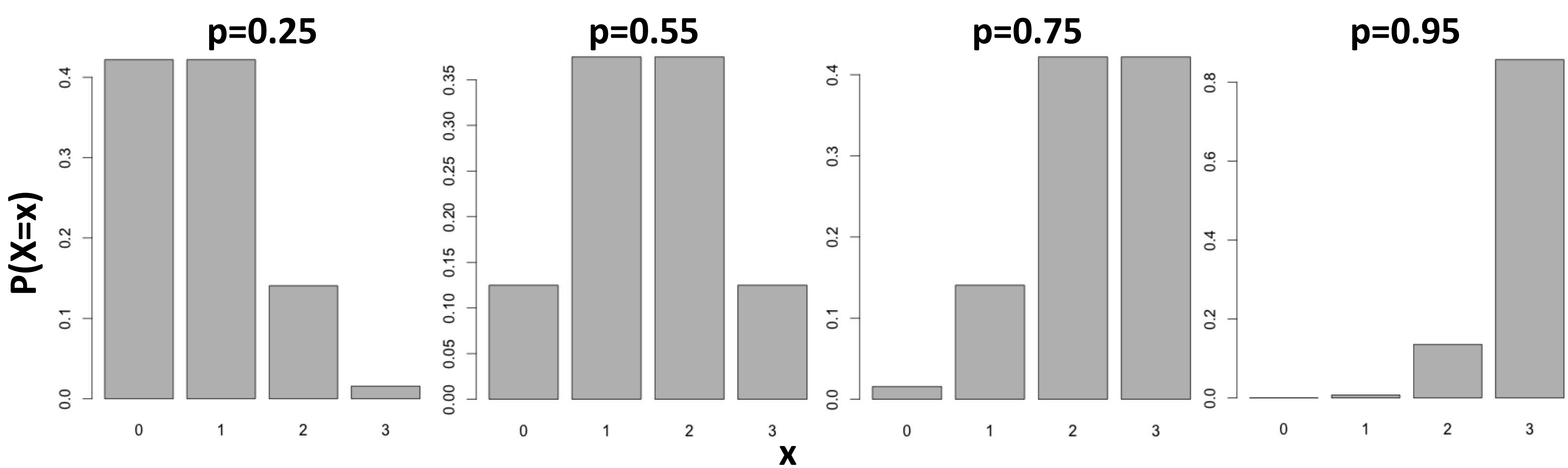
- We flip a fair coin 3 times.
  X = number of times we got heads. {0,1,2,3}
- What is the probability distribution of X?
  - The probability distribution describes the probability of seeing any particular value of X.
  - Probability distributions have *parameters*:
    we flipped a coin *3* times (size=3), and the coin is *fair* (p=0.5)
    different values of the parameters change the probability dist.

Binomial random variables

- We flip a *bent coin* [P(H)=p] 3 times.
  X = number of times we got heads. {0,1,2,3}
  What is the probability distribution of X?

- P(X=x) = sum(P(outcome)) for all outcomes where X=x
  P(X=0) = P(TTT) = 1 * (1-p)^3
  P(X=1) = P(TTH) + P(THT) + P(HTT) = 3 * p*(1-p)^2
  P(X=2) = P(THH) + P(HHT) + P(HTH) = 3 * p^2*(1-p)
  P(X=3) = P(HHH) = 1 * p^3

Binomial random variables



Binomial random variables

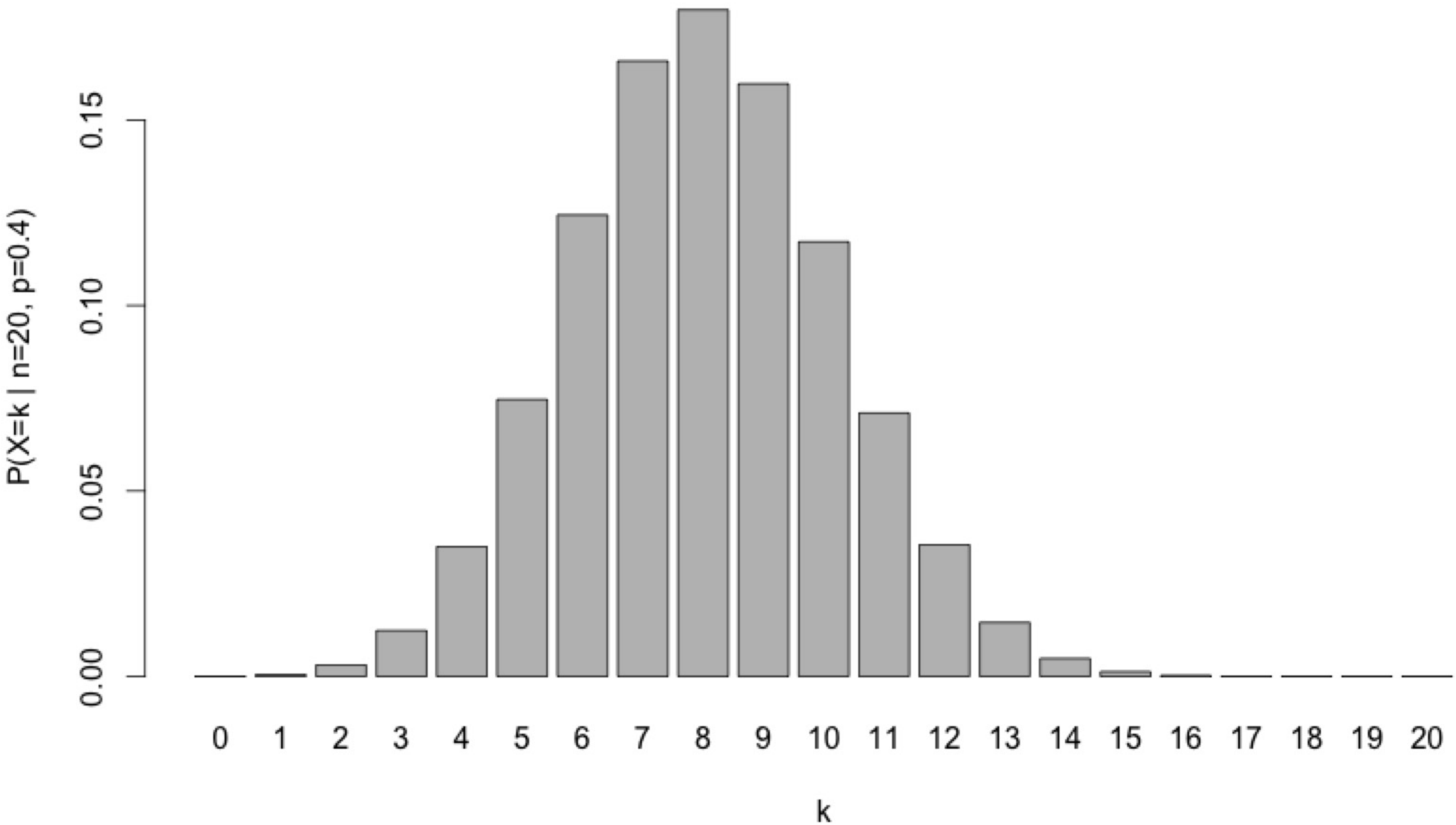
- We flip a bent coin (P(H)=p) *n* times.
  X = number of times we got heads. {0,1,2,3,...,n}
  What is the probability distribution of X?

- P(X=x) = sum(P(outcome)) for all outcomes where X=x
  P(X=0) = choose(n,0) * (1-p)^n
  P(X=1) = choose(n,1) * p * (1-p)^(n-1)
  P(X=2) = choose(n,2) * p^2 * (1-p)^(n-2)
  P(X=n) = choose(n,n) * p^n

- What we really want is a general expression:
  P(X=k) = choose(n,k) * p^k * (1-p)^(n-k)

Binomial random variables

X = number of heads from n flips of bent coin with P(H)=p



dbinom(x, size, p)

x: number of successes (e.g., number of heads)

size: number of attempts (e.g., number of coin flips)

p: probability of success (e.g., prob coin lands heads)

Returns: p(X=x) (where X is a binomial variable with parameters *size* and *p*

Probability distribution: density/mass:

- In R: d*

```
dbinom(x, …)
dbeta(x, …)
dnorm(x, …)
dgeom(x, …)
df(x, …)
dt(x, …)
```

- Each of these returns the probability *density or mass** of x, given their parameters.

Probability distribution function

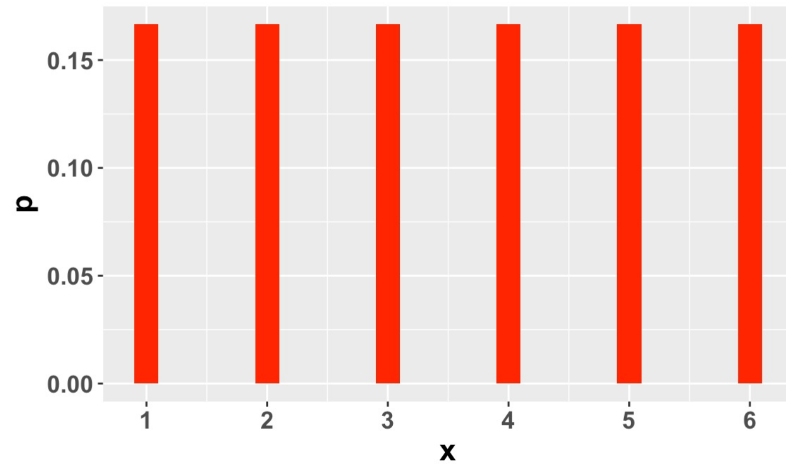
- Characterizes how probability is distributed over the possible values of the random variable, given the parameters.

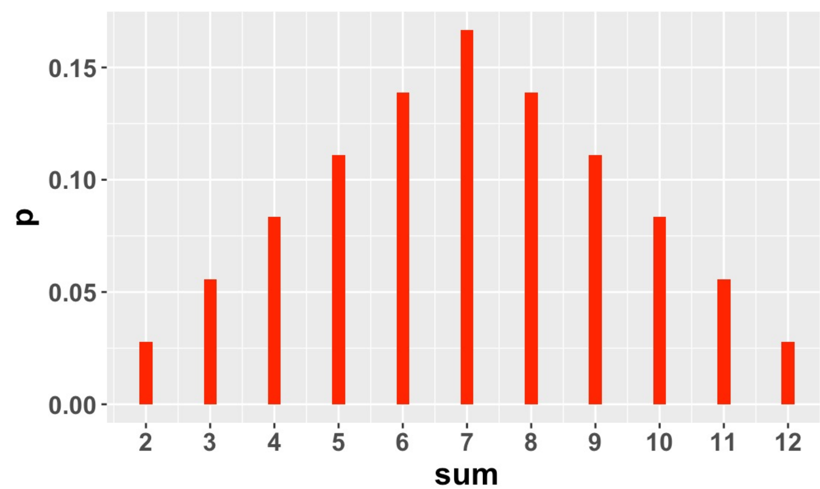
Discrete vs continuous

- Some are discrete, others continuous.
  - The outcome of a dice roll
  - The sum of two dice rolls
  - The number of rolls it takes to get a 6
  - The number of 5 or 6 rolls out of N attempts
  - The distance on the table the dice rolled when thrown
  - The decibel level of the groans of disappointment at a craps table in Vegas when the dice were rolled.
  - Etc.

Discrete RV: probability mass function

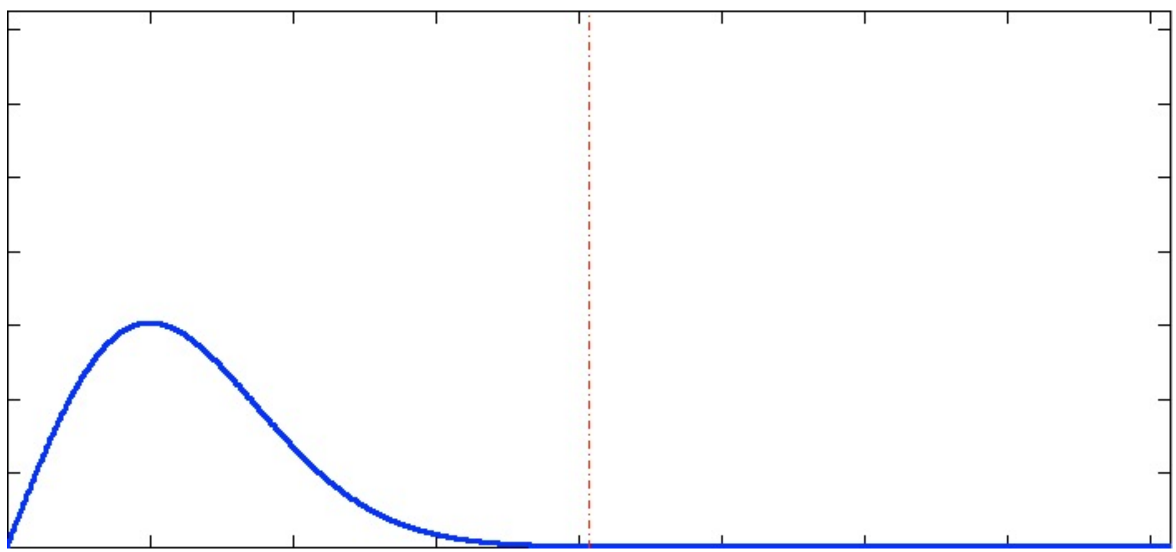
- Probability mass function (pmf):
  Each possible value has some non-zero probability.
  - The number of dots, the sum of two dice





Continuous RV: probability density

- **Probability density function** (pdf):
  Continuous variables have infinitely many values, each with infinitely little probability.
  Values only have "probability density"
  Intervals have probability mass
  (defined as 'area under the curve' of pdf in that interval)
  - The distance on the table the dice rolled when thrown



Probability mass vs density functions

- Discrete random variables
  - Every value has some definable probability mass.
  - If you sum Prob(x) for all values of x, you get 1.0
    e.g., `sum(dbinom(0:10, 10, 0.5)) = 1.0`
- Continuous random variables
  - Every value has 0 probability mass, because every value is infinitesimally precise. *Only ranges of values have probability mass*. Otherwise only prob. *density*
  - Probability density is defined as the derivative of the cumulative probability.
  - If you sum density(x) for "all" values of x, you get nonsense (nonsense that changes with how finely you slice x)

```
sum(dbeta(seq(0,1,by=0.01 ), 2, 2))=99.99
sum(dbeta(seq(0,1,by=0.001), 2, 2))=999.99
```

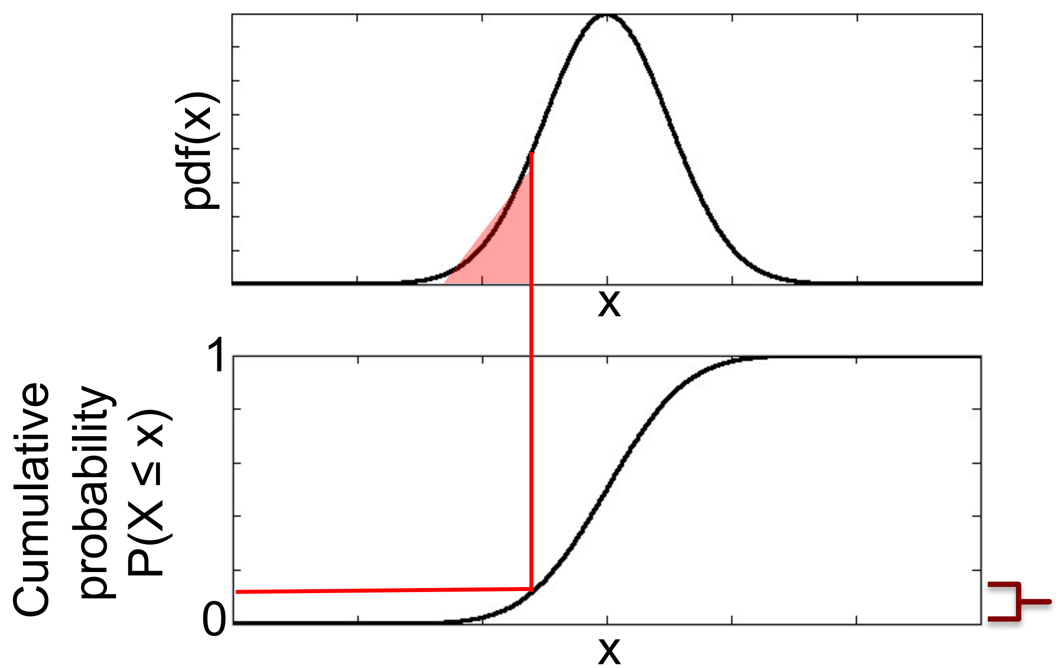
(of course, you could use numerical integration by summing density(x)*dx)

Probability

- Probability statements, philosophy
- Basic probability rules
- Joint, conditional, and Bayes' rule.
- *Random variables and probability distributions*
  - Sampling to generate random values.
  - Probability mass (pmf) and density (pdf) functions
  - **Cumulative probability density function (cdf)**
  - Quantile function (inverse cdf)
- Expectation, Variance, and their rules
- Central limit theorem and the normal distribution.

Cumulative distribution function (cdf)

- The integral /sum from -∞ to x of the pdf / pmf
- In other words, cdf(x) is the probability that a random variable will have a value less than or equal to x.



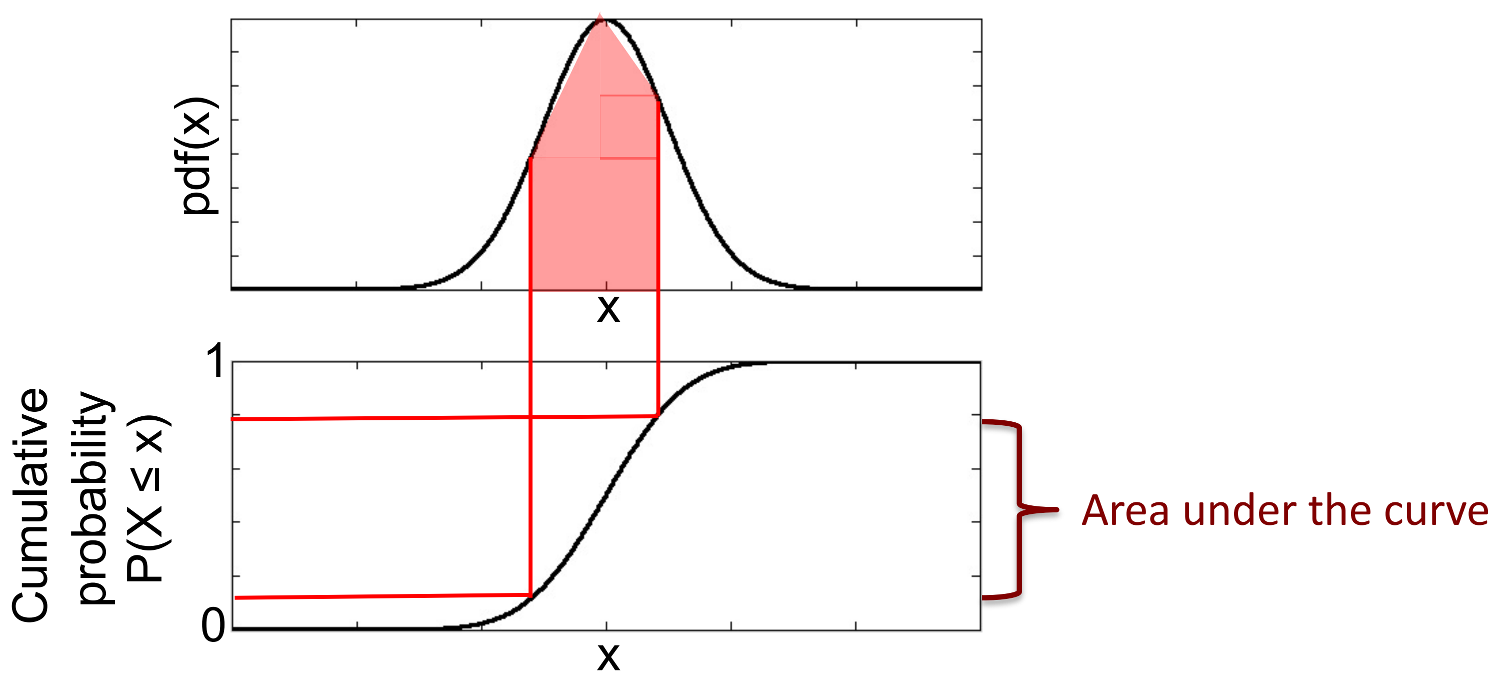
`dnorm(x, 0, 1)`

`pnorm(x, 0, 1)`

Area under the curve

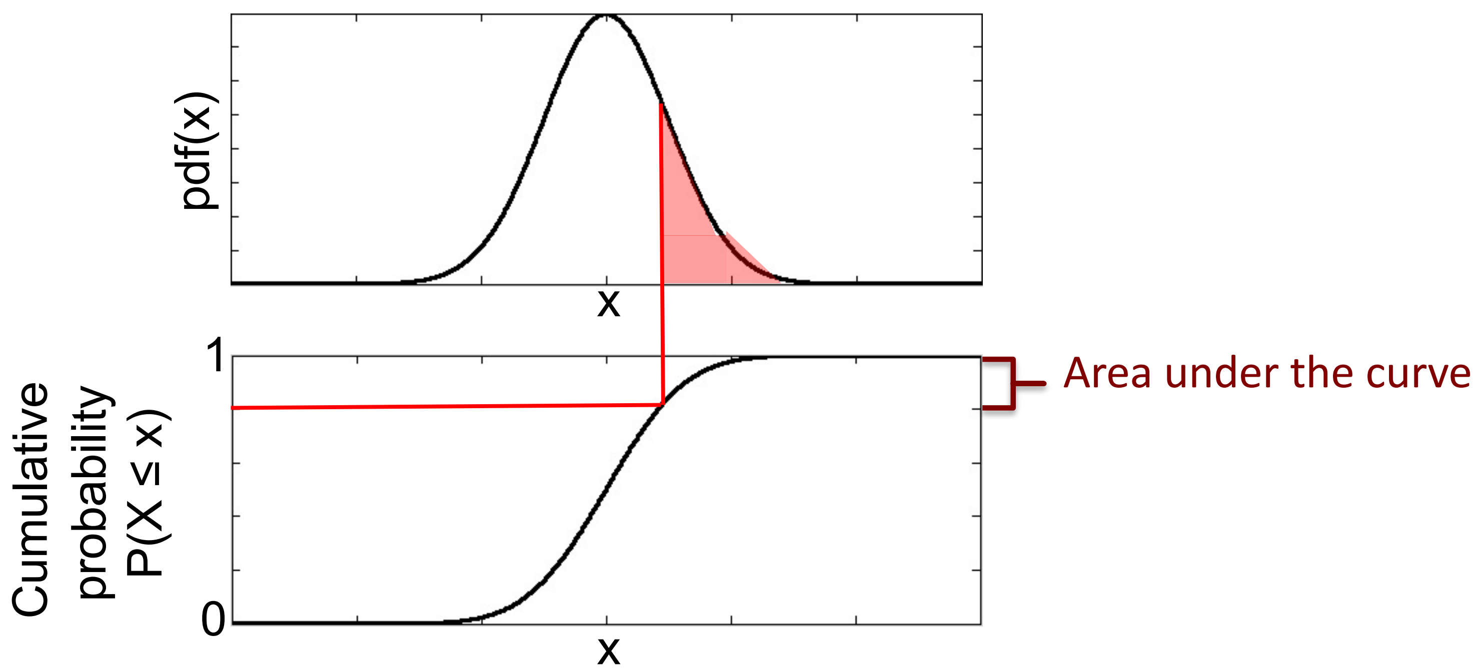
Cumulative distribution function (cdf)

- The integral /sum from -∞ to x of the pdf / pmf
- In other words, cdf(x) is the probability that a random variable will have a value less than or equal to x.



Cumulative distribution function (cdf)

- The integral /sum from -∞ to x of the pdf / pmf
- In other words, cdf(x) is the probability that a random variable will have a value less than or equal to x.



Cumulative probability

In general…		
	For discrete quant. var.	$P(\le x) = \sum_{y=-\infty}^{x} P(y)$
	For continuous quant. var.	$P(\le x) = \int_{-\infty}^{x} P(y)dy$

Gotcha with cdf for discrete variables.

- For continuous variables, $P(x) \approx 0$
- For discrete variables, $P(x) \geq 0$

Therefore…

- For continuous variables
  $P(X \geq x) = 1 - P(X \leq x)$
- For discrete variables
  $P(X \geq x) = 1 - P(X < x)$ **not** $1 - P(X \leq x)$

So if you want to evaluate the probability that a random variable will take on a value of x or higher…

…with a continuous variable you can calculate it as 1-CDF(X)

e.g., prob that a Normal(0,1) variable is 2 or m

```
1-pnorm(2, 0, 1)
```

…with a discrete variable over integers you must calculate it as 1-CDF(X-1) [**]

e.g., prob that we get 7 or more heads out of 10 coin flips is:

```
1-pbinom(7-1, size=10, prob=0.5)
```

pbinom(x, size, p)

x: number of successes (e.g., number of heads)

size: number of attempts (e.g., number of coin flips)

p: probability of success (e.g., prob coin lands heads)

Returns: $P(X \leq x)$ (where X is the random binomial variable).

E.g.,: probability that we get *x or fewer* heads.

pnorm(x, mean, sd)

x: value of variable (e.g., specific IQ score)

mean: average (e.g., mean IQ score is 100)

sd: standard deviation (e.g., IQ defined as sd=15)

Returns: $P(X \leq x)$ (where X is the random normal variable).

E.g.,: probability that an IQ score will be *x or lower*

Cumulative probability

- In R: p*

```
pbinom(x, …)
pbeta(x, …)
pnorm(x, …)
pgeom(x, …)
pf(x, …)
pt(x, …)
```

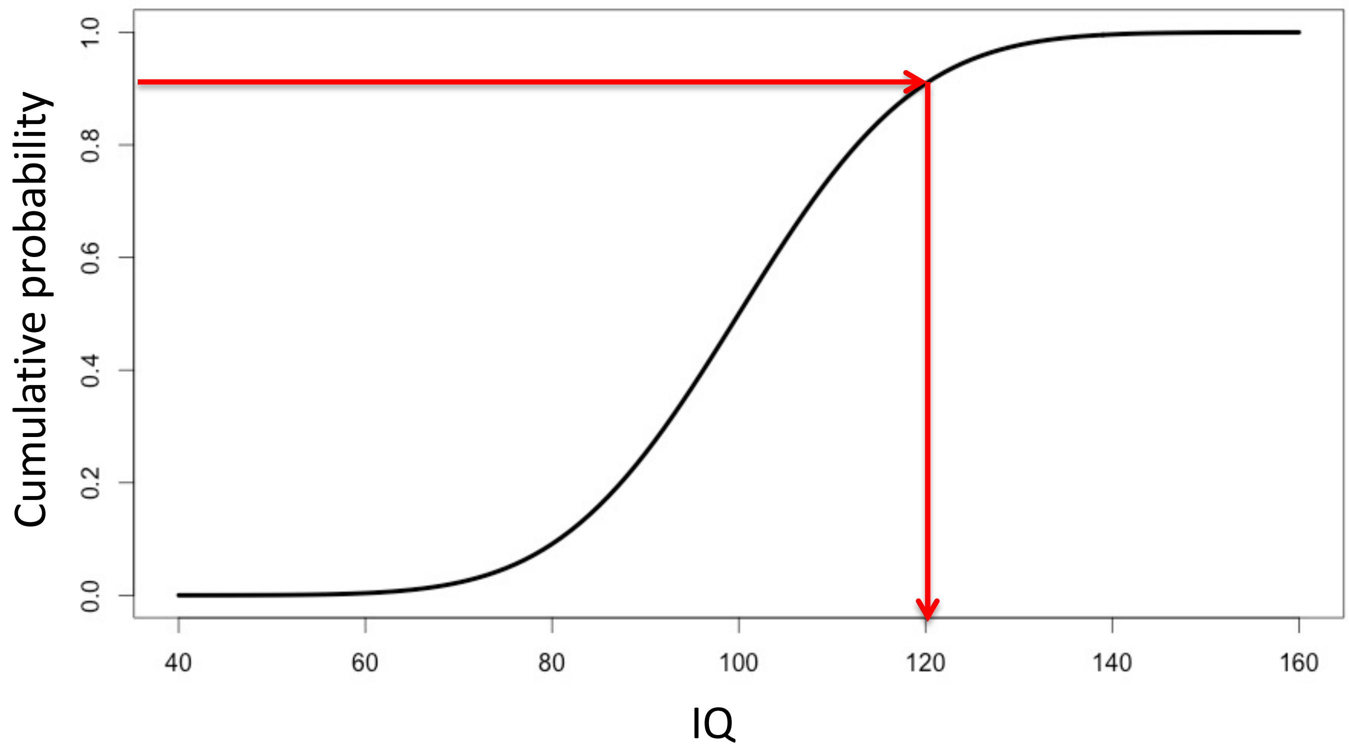
- Each of these returns the cumulative probability at x, given the parameters.

Probability

- Probability statements, philosophy
- Basic probability rules
- Joint, conditional, and Bayes' rule.
- *Random variables and probability distributions*
  - Sampling to generate random values.
  - Probability mass (pmf) and density (pdf) functions
  - Cumulative probability density function (cdf)
  - **Quantile function (inverse cdf)**
- Expectation, Variance, and their rules
- Central limit theorem and the normal distribution.

Quantile (inverse cdf) function

- What is the *value* such that the probability that this random variable is equal to or less than that value is *q*?



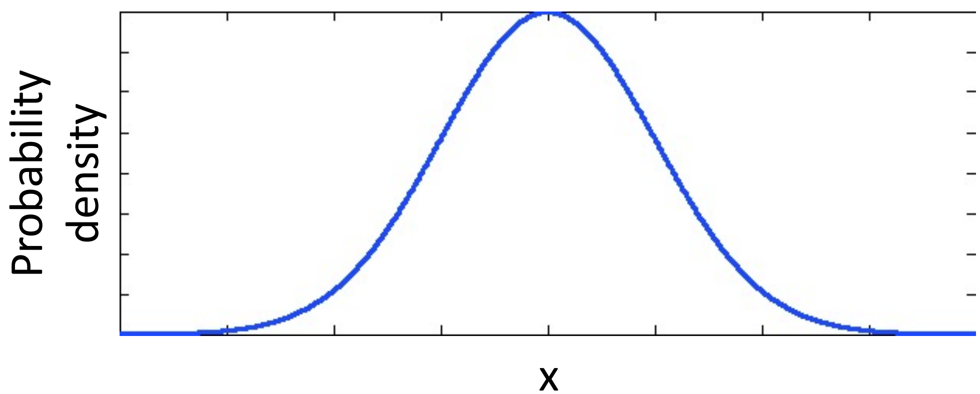
What is the IQ such that 90% of all people have a lower IQ than that?

120

Note: for discrete variables there often isn't an exact quantile: it falls between two values. Different methods exist for interpolating, or not. We will use the default R quantile functions as the gold standard.

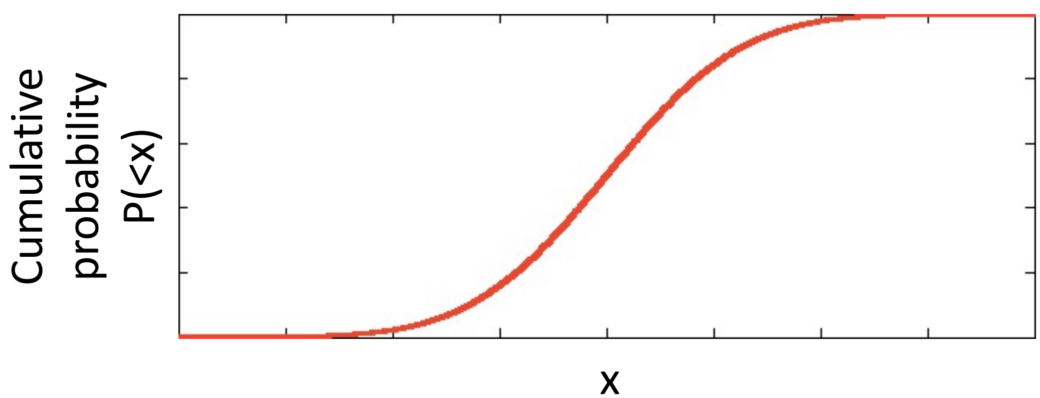
Quantile function (inverse cdf)

Normal PDF



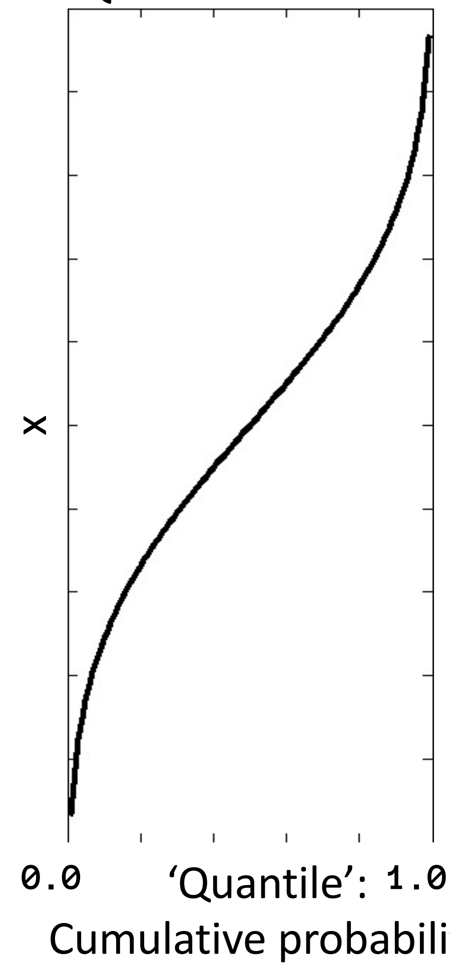

```
plot(x, dnorm(x, 0, 1))
```

Normal CDF




```
plot(x, pnorm(x, 0, 1))
```

Normal iCDF 'Quantile' function




```
plot(p, qnorm(p, 0, 1))
```

CDF and Quantile

- Cumulative distribution function (cdf):
  p = cdf(x)

```
pnorm(120, mean=100, sd=15)
```

  'an IQ of 120 puts me in what quantile (percentile) of the IQ distribution?'

- Quantile function (icdf):
  x = icdf(p)

```
qnorm(0.91, mean=100, sd=15)
```

  'what would my IQ have to be to be in the 91st percentile?'

Variants of a quantile

- **Median:** 0.50th quantile.
  (although "median" usually refers to a sample statistic, not a property of a distribution)
- **Percentile**: 1st: 0.01; 2nd: 0.02; ...; 98th: 0.98, etc.
- **Quartile**: 1st: 0.25; 2nd: 0.5; 3rd: 0.75
  'Interquartile range': distance between 1st and 3rd quartile.
- **Quintile:** 1st: 0.2; 2nd: 0.4; ...; 4th: 0.8
- **Decile:** 1st: 0.1; 2nd: 0.2; ...; 9th: 0.9.

qnorm(p, mean, sd)

p: Cumulative probability (e.g., IQ percentile)

mean: average (e.g., mean IQ score is 100)

sd: standard deviation (e.g., IQ defined as sd=15)

Returns: x such that $P(X \leq x) = p$
(where X is the random normal variable).

E.g.,: the IQ score that will be at the p^{th} quantile;

IQ score such that *p* of all IQ scores are less than it.

(quantile functions for discrete RVs are tricky)

Quantile function

- In R: q*

```
qbinom(p, …)
qbeta(p, …)
qnorm(p, …)
qgeom(p, …)
qf(p, …)
qt(p, …)
```

- Each of these returns the x such that the cumulative probability at x is *p*, given the parameters.

Probability

- Probability statements, philosophy
- Basic probability rules
- Joint, conditional, and Bayes' rule.
- *Random variables and probability distributions*
  - Sampling to generate random values.
  - Probability mass (pmf) and density (pdf) functions
  - Cumulative probability density function (cdf)
  - Quantile function (inverse cdf)
- **Expectation, Variance, and their rules**
- Central limit theorem and the normal distribution.

Expectation and moments of RVs

- BEWARE:
  Most moments have a sample statistic counterpart (e.g., mean, variance), but there is a difference between calculating the *sample statistic* and calculating the *expectation of a function of a random variable.*

"Expectation" / "Expected value"

Expected value of a random variable is its mean:
the sum of each possible value weighted by its probability

$$\boxed{E[X]} = \boxed{x_1 P(X = x_1) + x_2 P(X = x_2) + x_3 P(X = x_3) + \ldots + x_k P(X = x_k)}$$

Expected value of X.
"Mean[X]"

Sum of all values of x weighted by their probability mass

Note: relationship to sample statistic via Monte Carlo theorem

Expected value

Expected value of a random variable is its mean:
the sum of each possible value weighted by its probability

- Sum for discrete variables; integral for continuous.

$$E[X] = \boxed{\sum_{x \in X}} xP(X = x)$$

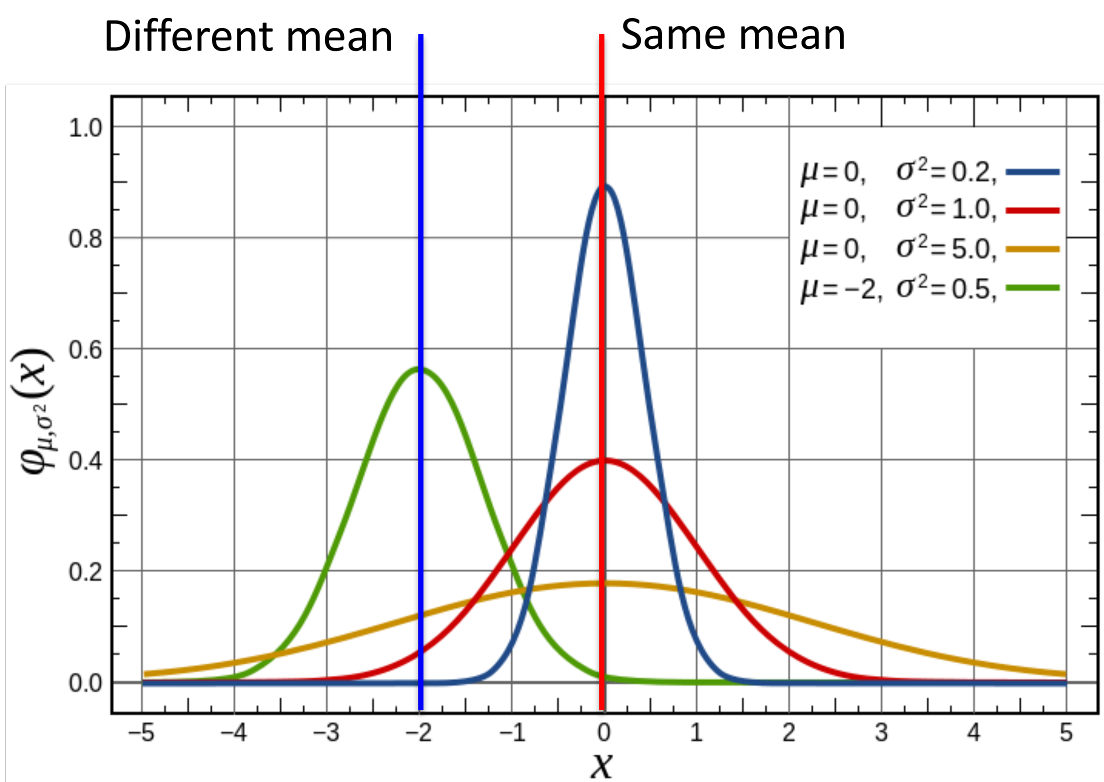
Set notation for sum over all the possible values of X.

$$E[X] = \int_{-\infty}^{\infty} f(x)x\boxed{dx}$$

The dx in the integral corresponds to the infinitely small interval around x. By taking this interval in account f(x)dx yields a probability mass.

Expectation and moments

- Mean: $\mu_X = Mean[X] = E[X]$
  the 'location' of the random variable: where is the variable centered? Where is the center of mass?



Note that not all distributions are symmetric, so the mean is not always the mode.

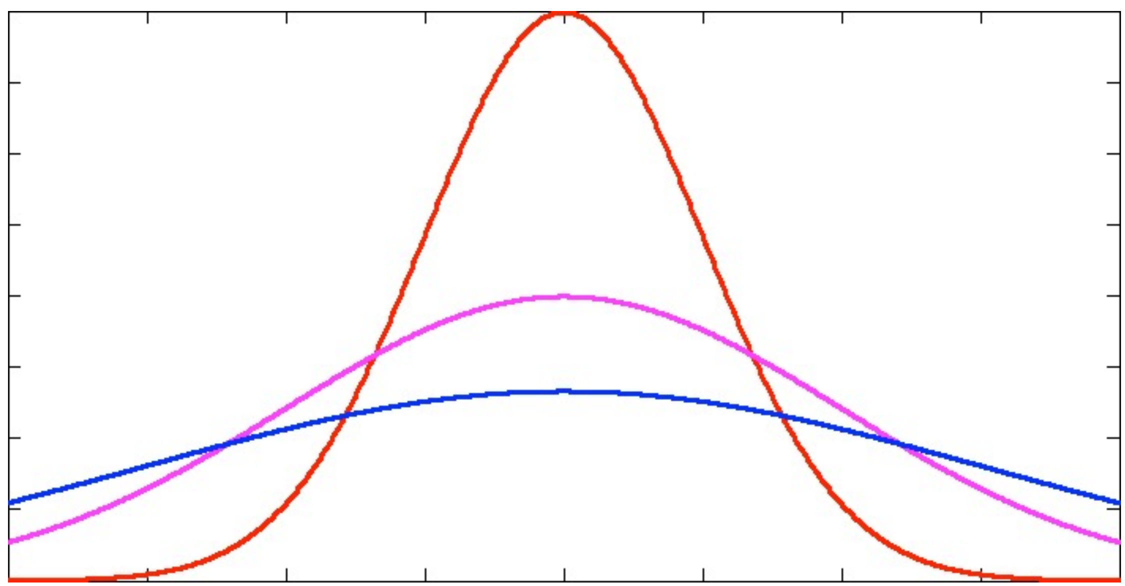
Expectation and moments

- Mean: $\mu_X = Mean[X] = E[X]$

- Note:
  here we are using 'mu_X' rather than 'x-bar' to denote the mean because we are talking about the mean of a random variable with some known probability distribution, ***not the sample mean***, which is a statistic of some data.
- Typically we use the sample mean to estimate mu – the mean of the random variable which we sampled.

Expectation and moments

- Variance: $\sigma_X^2 = Variance[X] = E[(X - E[X])^2] = E[X^2] - E[X$
  The ‘scale’ of the distribution: how spread out is the probability over possible values of x?

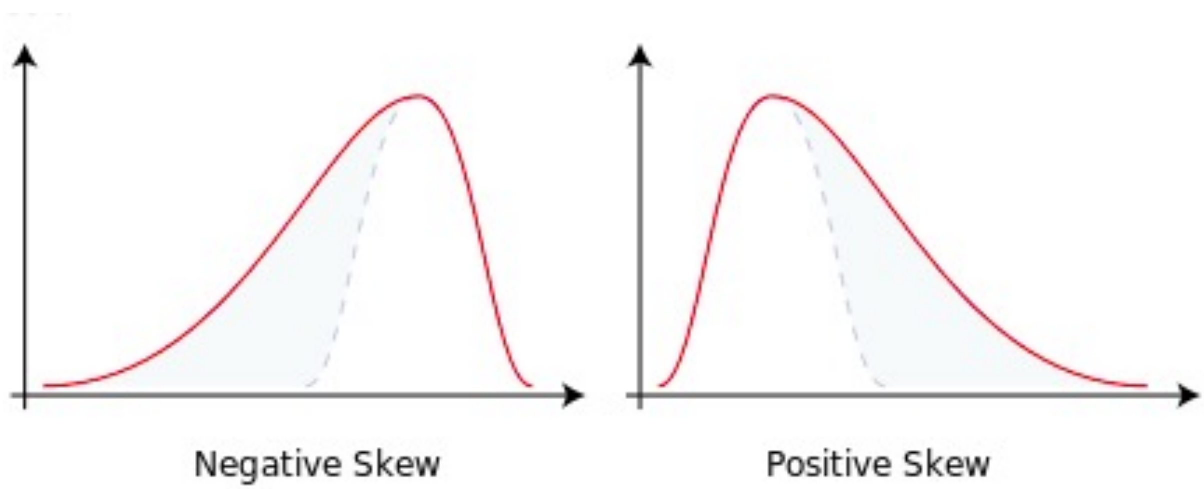


Expectation and moments

- Skewness:

$$Skew[X] = E\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^3\right]$$

is the distribution asymmetric? If so, is the negative or positive tail heavier?



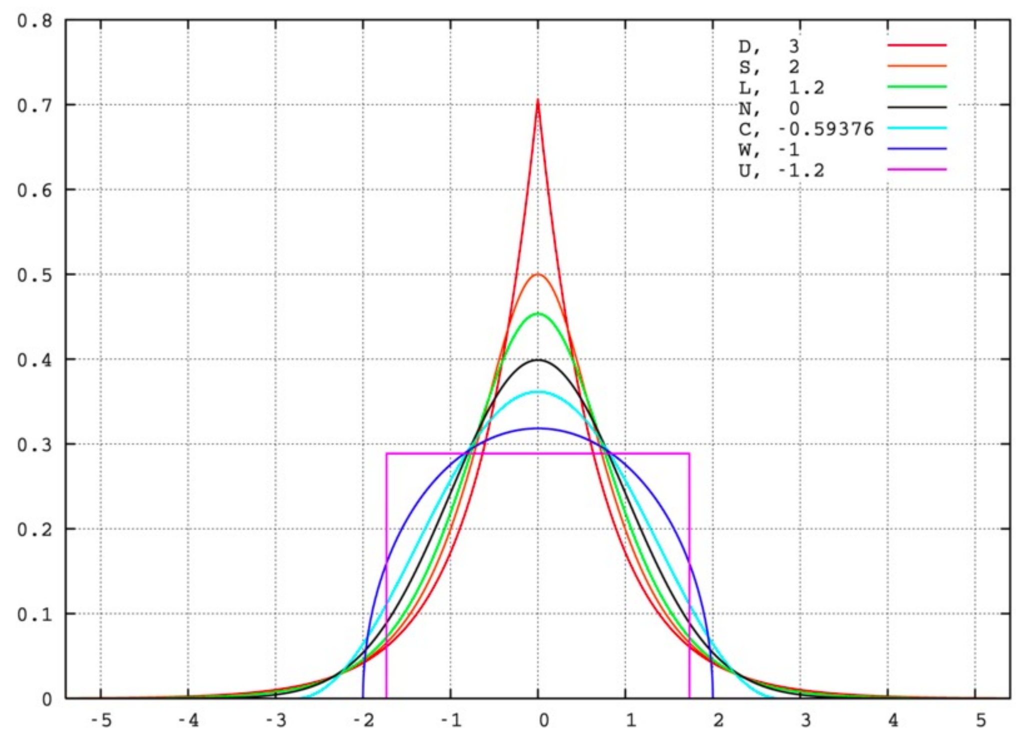
Expectation and moments

- (Excess) Kurtosis:

$$Kurtosis[X] = E\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^4\right] \boxed{-3}$$

is the distribution peakier with heavier tails (high – positive) or squat with shorter tails (low – negative)?

Subtracting 3 (the kurtosis of a normal distribution) makes this into 'Excess' Kurtosis relative to the normal



Expectation and moments

- Mean: $$\mu_X = Mean[X] = E[X]$$

- Variance: $$\sigma_X^2 = Variance[X] = E[(X - E[X])^2]$$

- Skewness: $$Skew[X] = E\left[\left(\frac{x - \mu_X}{\sigma_X}\right)^3\right]$$

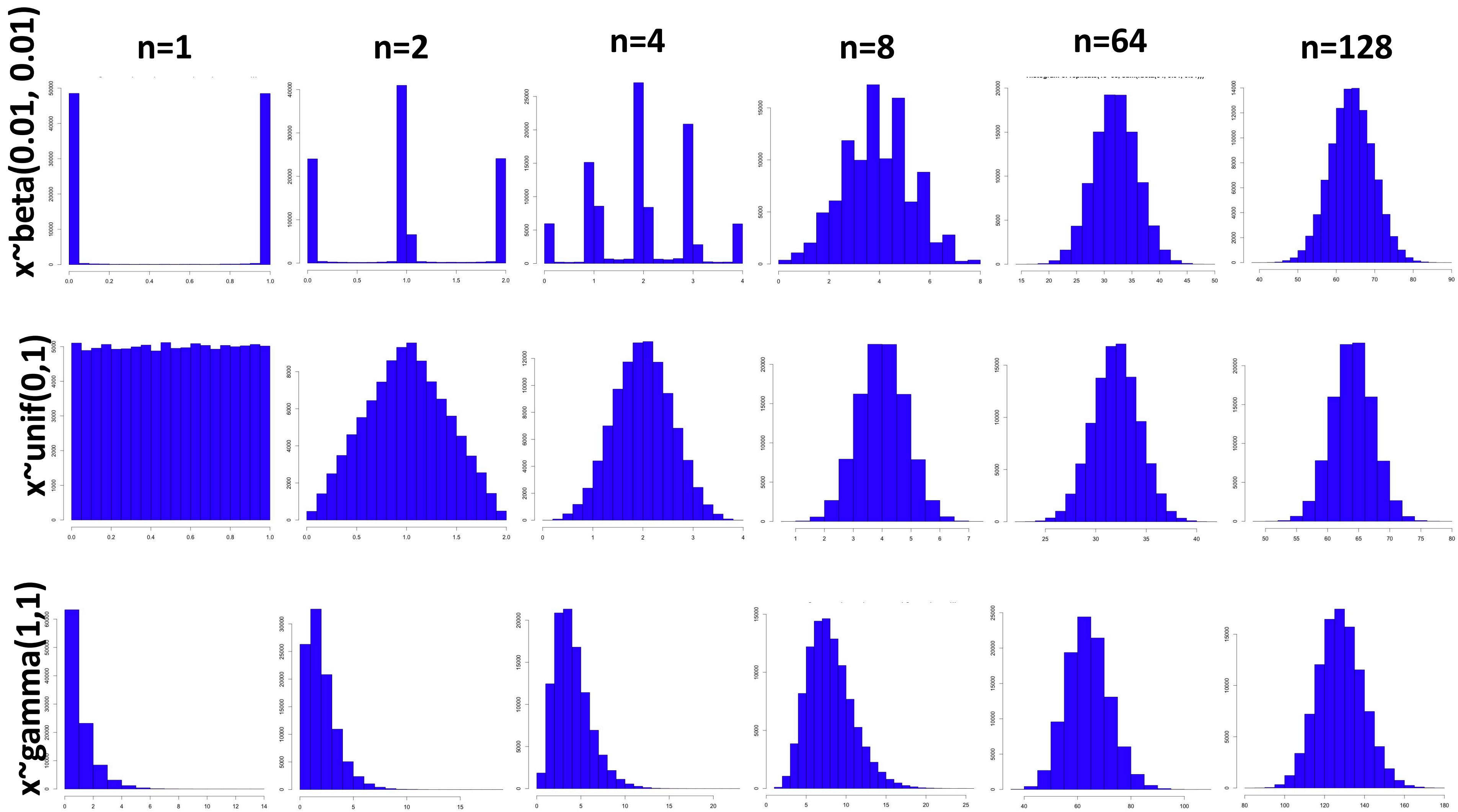
- (Excess) Kurtosis: $$Kurtosis[X] = E\left[\left(\frac{x - \mu_X}{\sigma_X}\right)^4\right] - 3$$

Note: all of these are the expected value of some f(x) weighted by P(x): sum(f(x)*p(x))
For Mean[X] f(x) = x
For Variance[X] f(x) = (x-Mean[X])^2
We can calculate the expectation for any* f(x) of a random variable.

Probability

- Probability statements, philosophy
- Basic probability rules
- Joint, conditional, and Bayes' rule.
- *Random variables and probability distributions*
  - Sampling to generate random values.
  - Probability mass (pmf) and density (pdf) functions
  - Cumulative probability density function (cdf)
  - Quantile function (inverse cdf)
- Expectation, Variance, and their rules
- **Central limit theorem and the normal distribution.**

Distribution of the sum of *n* iid RVs

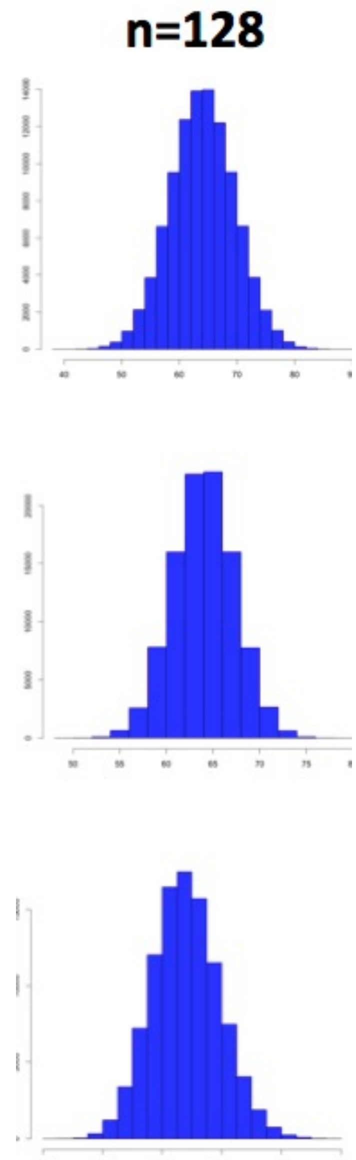



```
replicate(100000, sum(rgamma(n,1,1))
```

Central limit theorem

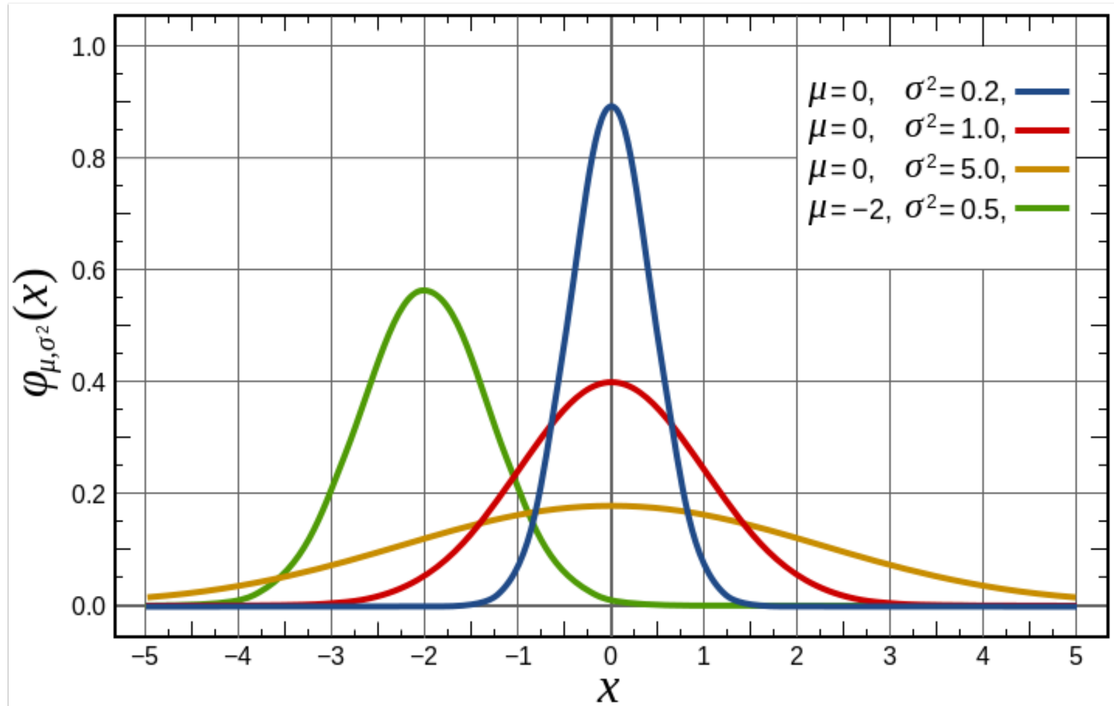
- The sum of *n* i.i.d. random variables is Normally distributed if *n* is big enough*

- Many real-world variables can be thought of as the sum of lots of independent and roughly identically distributed, contributing factors, so we often treat our measures as having a Normal distribution, but this should be verified.



Normal Distribution

It has two parameters:
"location" (mean; mu)
"scale" (sd or var)



In R for a Normal distribution with mean M and sd S

Probability density at x = dnorm(x, M, S)

Cumulative probability at x = pnorm(x, M, S)

Quantile function for p = qnorm(p, M, S)

Probability

- <u>Probability statements, philosophy</u>
- <u>Basic probability rules</u>
- <u>Joint, conditional, and Bayes' rule.</u>
- <u>*Random variables and probability distributions*</u>
  - Sampling to generate random values.
  - Probability mass (pmf) and density (pdf) functions
  - Cumulative probability density function (cdf)
  - Quantile function (inverse cdf)
- Expectation, Variance, and their rules
- Central limit theorem and the normal distribution.

Some typical probability questions.

- When flipping a fair coin, what is the probability that the first occurrence of heads will be on the 5th flip?

- If you run 20 independent, tests on truly null data, each with a false-positive rate of 0.05, what's the probability that you will get at least 1 false positive (the familywise false-positive rate)?
  - What would the *per-test* false-positive rate have to be for the *familywise* false positive rate –
    P(at least 1 false-positive) -- to be 0.05?

Binomial random variables

- We flip a *bent coin* [P(H)=p] *n* times.
  X = number of times we got heads. {0,1,2,3,...,n}
  What is the probability distribution of X?

- P(X=x) = sum(P(outcome)) for all outcomes where X=x
  P(X=0) = 1 * (1-p)^n
  P(X=1) = # * p * (1-p)^(n-1)
  P(X=2) = # * p^2 * (1-p)^(n-2)
  P(X=n) = 1 * p^n

- What we really want is a general expression:
  P(X=k) = # * p^k * (1-p)^(n-k)

- What are the #s?
  How many different ways are there to get *k* / *n* heads?

Binomial coefficient

- How many ways are there to get k heads in n flips?

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

- ‘N choose K’

```
factorial(5)/factorial(3)/factorial(2)
```

```
[1] 10
```

```
choose(5,3)
```

```
[1] 10
```

Binomial random variables

- We flip a bent coin (P(H)=p) *n* times.
  X = number of times we got heads. {0,1,2,…,n}

$$\boxed{P\left(X=k \mid n,p\right)} = \boxed{\binom{n}{k}}\boxed{p^{k}(1-p)^{n-k}}$$

Probability that binomial random variable have k successes. Given that there are n attempts, and each one has probability p of being a 'success'

Binomial coefficient: How many unique outcomes are there with k successes out of n attempts?

Probability of one specific outcome that has k successes out of n attempts

Binomial probability *mass* function

This is the Binomial probability mass function.
It tells us the probability that (Binomial) random variable X will take on a given value (*k*) given two parameters (*n*, *p*)
Note that these parameters have a *meaning*:
n: number of coin flips
p: bias of the coin (probability of heads/success)

Binomial random variables

- We flip a bent coin (P(H)=p) *n* times.
  X = number of times we got heads. {0,1,2,...,n}

```
n=20
p=0.4
k=7
choose(n,k)*p^k*(1-p)^(n-k)
```

```
[1] 0.1797
```

```
dbinom(k,n,p)
```

```
[1] 0.1797
```

```
x=0:20
```

```
[21] 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
```

```
choose(n,x)*p^x*(1-p)^(n-x)
```

```
[21] 3.656158e-05 4.874878e-04 3.087423e-03 …
```

```
dbinom(x,n,p)
```

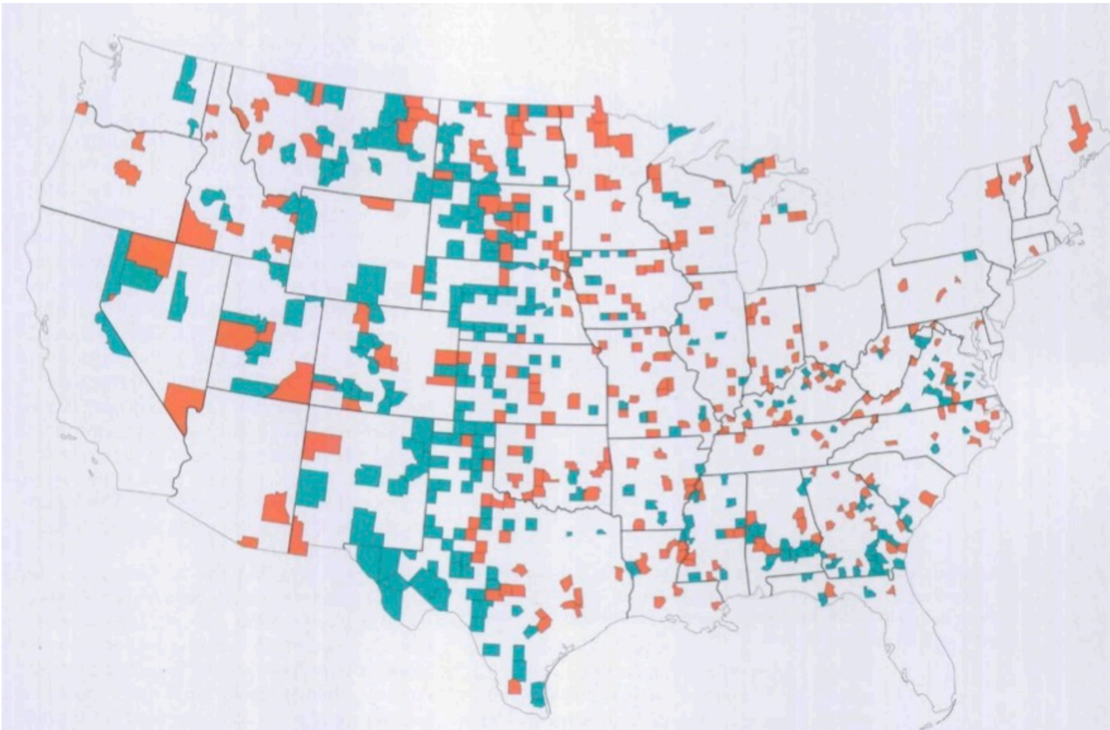
```
[21] 3.656158e-05 4.874878e-04 3.087423e-03 …
```

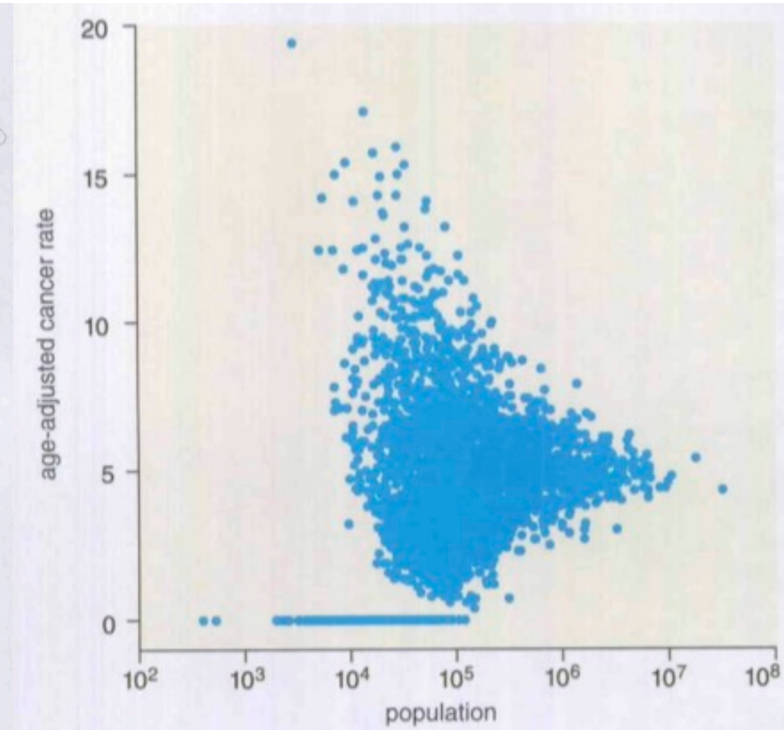
Binomial questions

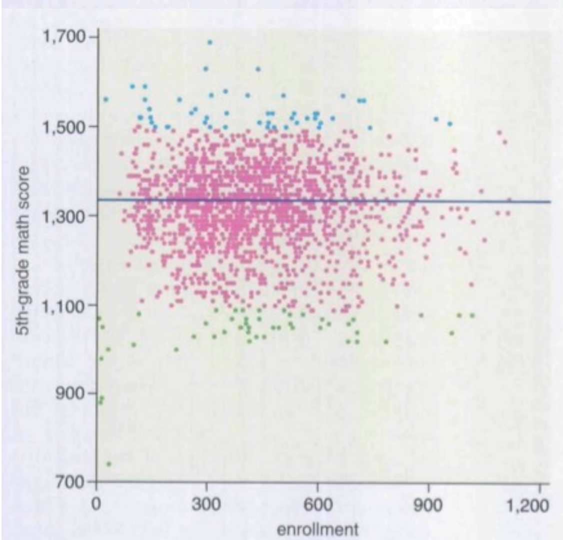
Use vectors, dbinom(), and sum() in R to make these calculations!

- Generally, 51.2% of all (US) births are male.
  A hospital has 10 births in one day.
  What is the probability that...
  - exactly 6 of them will be male?
  - 7 or more of them will be female?
  - The probability that the proportion of male births will be abnormal
    i.e., either abnormally high (>75%) **or** abnormally low (<25%)
  - What is the probability of an abnormal proportion of male births in a hospital that has 100 births in one day?

Small samples and variability







Extreme events are way more likely in small samples!

"The Most Dangerous Equation" by Howard Wainer.

"Ignorance of how sample size affects statistical variation has created havoc for nearly a millennium"

Binomial questions with rbinom()

Use rbinom() to make these calculations!

- Generally, 51.2% of all (US) births are male.
  A hospital has 10 births in one day.
  What is the probability that...
  - exactly 6 of them will be male?
  - 7 or more of them will be female?
  - The proportion of male births will be abnormal
    i.e., either abnormally high (>75%) **or** abnormally low (<25%)

  - What is the probability of an abnormally high proportion of female births in a hospital that has 100 births in one day?

Cumulative probability questions

use pnorm(), pbinom() in R

- IQ normally distributed with a mean=100 and sd=15.
  What is the probability that a given person has an IQ…
  - Less than 120?
  - Greater than 145?
  - Between 90 and 110?
- Test scores on a 25 item quiz are binomially distributed with n=25, p=0.7.
  What is the probability that a given person's score is…
  - Less than 17?
  - Greater than 20?
  - Between 15 and 20?

Quantile questions

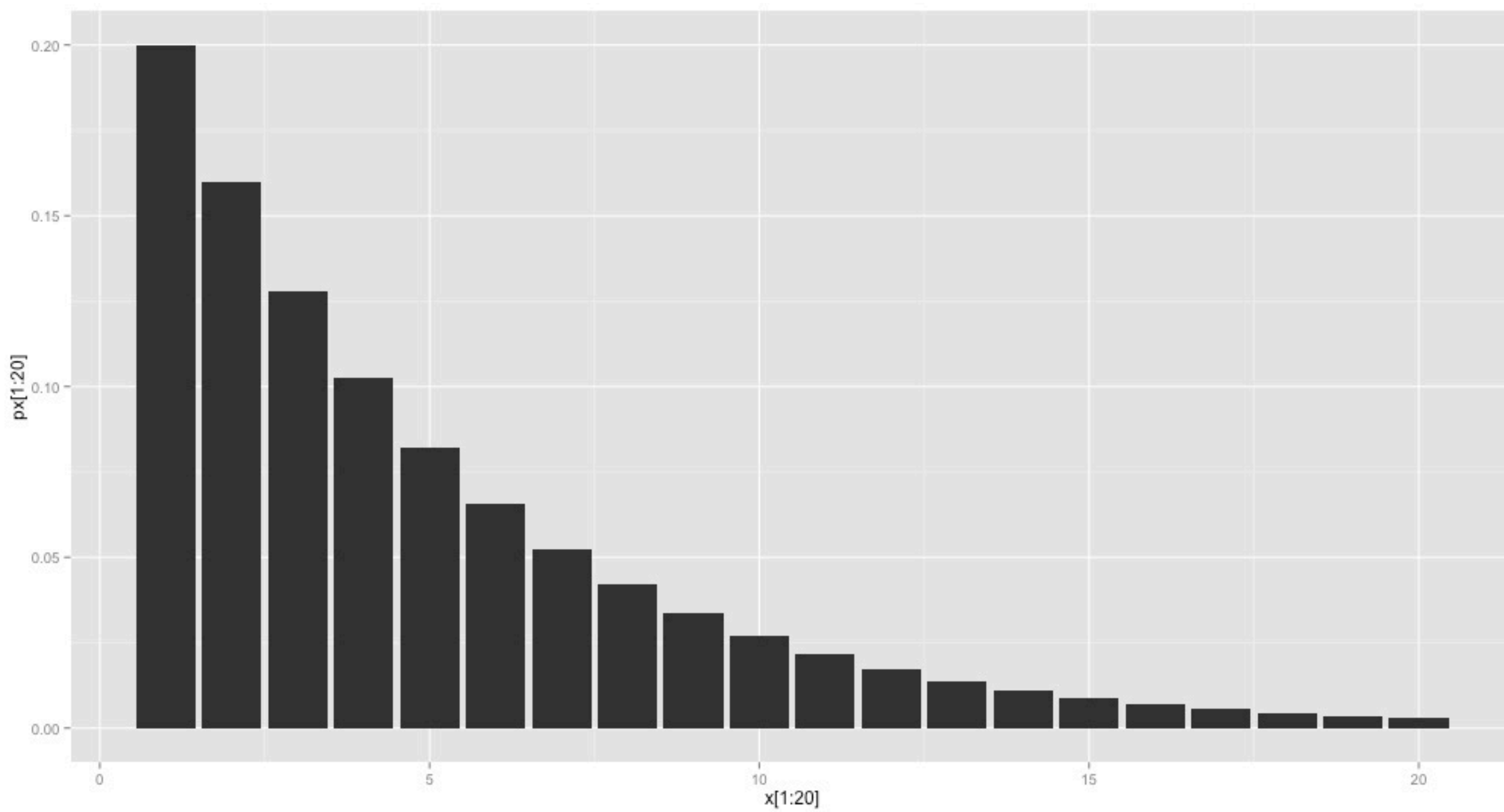
Use qnorm(), qbinom() in R

- IQ normally distributed with a mean=100 and sd=15.
  - What would your IQ score have to be to join MENSA (98th percentile of IQ distribution)?
  - What is the interquartile range of IQ?
  - The Prometheus society accepts only the top 1/30000th of the IQ distribution. How much higher is the IQ cutoff for Prometheus membership as compared to MENSA?
- Test scores on a 25 item quiz are binomially distributed with n=25, p=0.7.
  - What score would put you in the 90th percentile?
  - What is the interquartile range for these scores?
  - What is the median score?

Calculating expectation

Let's say I flip a bent coin, that comes up heads with probability 0.2. What is the expected number of flips until I see my first heads?

Probability distribution of X: the flip on which I see the first heads.



Calculating expectation

Let's say I flip a bent coin, that comes up heads with probability 0.2. What is the expected number of flips until I see my first heads?

Random variable: the position in the sequence when the first heads comes up.
Possible values of the random variable x
(note, possible values are integers from 1 to infinity, but here 10K is more than big enough to include all possibilities with any considerable probability)

```
x = 1:10000
```

```
[10000] 1, 2, 3, 4, …, 9999, 10000
```

Parameter of the probability distribution of this random variable: p(heads)

```
p.h = 0.2
```

```
[1] 0.2
```

Probability of every value of x

```
px = (1-p.h)^(x-1)*p.h
```

```
[10000] 0.2, 0.16, 0.128, 0.1024, …, 0, 0
```

Expected value of x (sum of all values of x weighted by their probabilities)

```
M.x = sum(x*px)
```

```
[1] 5
```

So, on average, it will take 5 coin flips to see our first heads…

Calculating expectation

Let's say I flip a bent coin, that comes up heads with probability 0.2. What is the expected number of flips until I see my first heads?

Random variable: the position in the sequence when the first heads comes up.
Possible values of the random variable x
(note, possible values are integers from 1 to infinity, but here 10K is more than big enough to include all possibilities with any considerable probability)

```
x = 1:10000                                   [10000] 1, 2, 3, 4, …, 9999, 10000
```

Parameter of the probability distribution of this random variable: p(heads)

```
p.h = 0.2                                     [1] 0.2
```

Probability of every value of x

```
px = (1-p.h)^(x-1)*p.h
                              [10000] 0.2, 0.16, 0.128, 0.1024, …, 0, 0
```

Expected value of x (sum of all values of x weighted by their probabilities)

```
M.x = sum(x*px)                               [1] 5
```

Expected variance of x (sum of all values of (x-M.x)^2 weighted by their probabilities)

```
V.x = sum( (x-M.x)^2 * px )                   [1] 20
```

Calculating expectation

Let's say I flip a bent coin, that comes up heads with probability 0.2. What is the expected number of flips until I see my first heads?

Possible values of the random variable x

```
x = 1:10000                    [10000] 1, 2, 3, 4, …, 9999, 10000
p.h = 0.2                      [1] 0.2
```

Probability of every value of x

```
px = (1-p.h)^(x-1)*p.h         [10000] 0.2, 0.16, 0.128, …, 0, 0
```

Expected value of x (sum of all values of x weighted by their probabilities)

```
M.x = sum(x*px)                [1] 5
```

Expected variance of x (sum of all values of (x-E.x)^2 weighted by their probabilities)

```
V.x = sum( (x-M.x)^2 * px )    [1] 20
```

Expected skew of x (sum of all values of Z(x)^3 weighted by their probabilities)

```
Zx = function(x)( (x-M.x) / sqrt(V.x) )
Skew.x = sum( Zx(x)^3 * px )   [1] 2.02
```

Calculating expectation

Let's say I flip a bent coin, that comes up heads with probability 0.2. What is the expected number of flips until I see my first heads?

Possible values of the random variable x

```
x = 1:10000                                [10000] 1, 2, 3, 4, …, 9999, 10000
p.h = 0.2                                  [1] 0.2
```

Probability of every value of x

```
px = (1-p.h)^(x-1)*p.h                     [10000] 0.2, 0.16, 0.128, …, 0, 0
```

Expected value of x (sum of all values of x weighted by their probabilities)

```
M.x = sum(x*px)                            [1] 5
```

Expected variance of x (sum of all values of (x-E.x)^2 weighted by their probabilities)

```
V.x = sum( (x-M.x)^2 * px )                [1] 20
```

Expected skew of x (sum of all values of Z(x)^3 weighted by their probabilities)

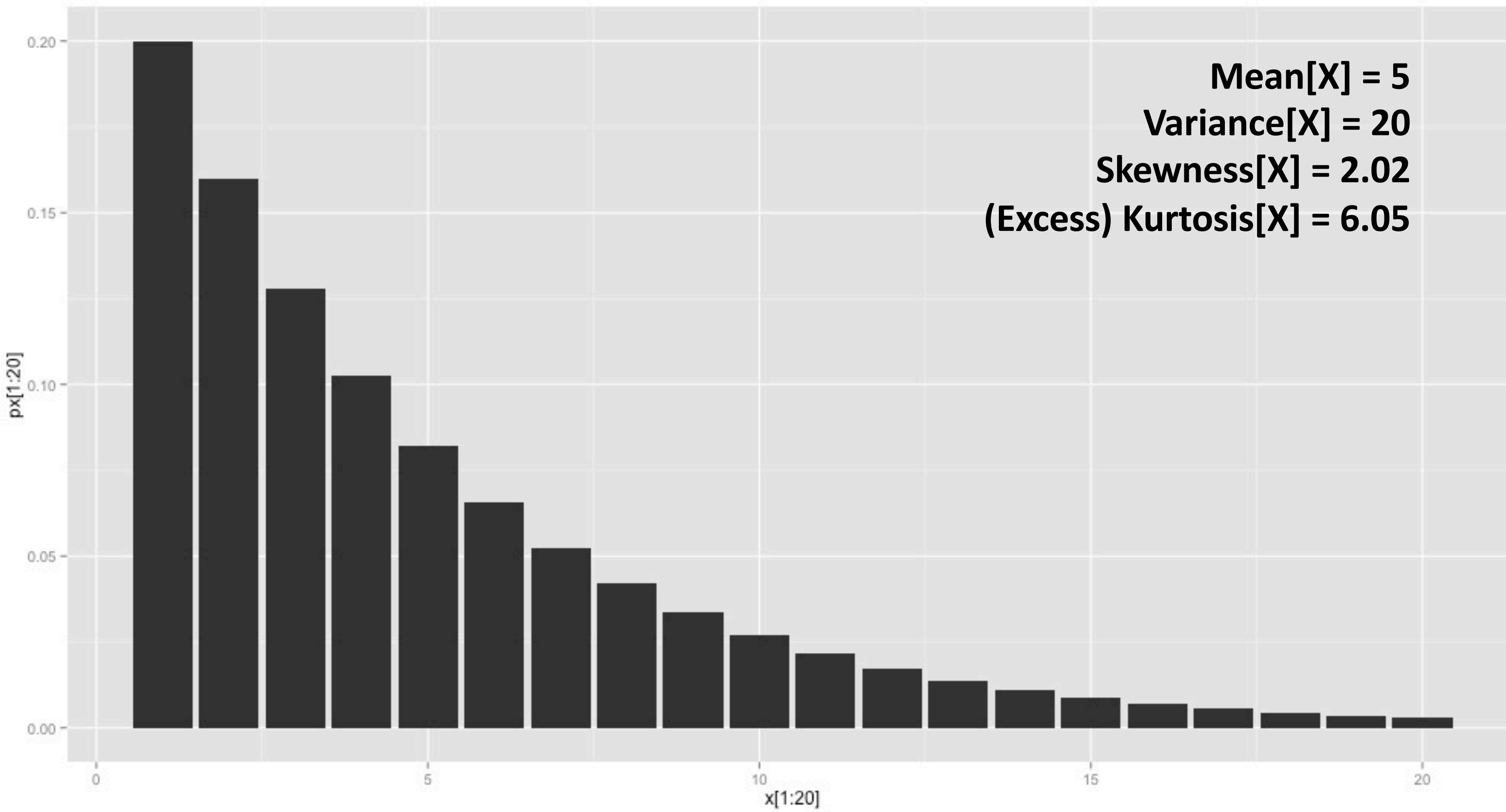
```
Zx = function(x)( (x-M.x) / sqrt(V.x) )
Skew.x = sum( Zx(x)^3 * px )               [1] 2.02
```

Expected (Excess) Kurtosis of x (sum of all values of Z(x)^4 weighted by their probabilities)

```
Kurt.x = sum( Zx(x)^4 * px ) - 3           [1] 6.05
```

Calculating expectation

Let's say I flip a bent coin, that comes up heads with probability 0.2. What is the expected number of flips until I see my first heads?



Expectation questions

1) Scores (X) on a 25 item quiz are distributed as a Binomial with $n=25$, $p=0.7$
   - What is the expected value of X?
   - Variance of X?
   - Skew of X?
   - Kurtosis of X?
2) Scores on another quiz (Y) have Mean[Y]=15, Var[Y]=16 (and are independent with scores on the first quiz). I add both scores up to get the final score, Z.
   - What is Mean[Z]?
   - What is Var[Z]?

Properties of mean and variance

- Properties:

$$Mean[aX+b] = a \cdot Mean[X] + b$$

$$Variance[aX+b] = a^2 \cdot Variance[X]$$

$$Mean[X+Y] = Mean[X] + Mean[Y]$$

$$Variance[X+Y] = Variance[X] + Variance[Y]$$

Nefarious question:

X has mean 5, variance 6;
Y has mean -4, variance 10.
Z = 5X-3Y+10.
What are the mean and variance of Z?

This is useful to know because the various cryptic equations for e.g., a t-statistic for various kinds of t-tests are derived from these sorts of calculations.

Not a proof of central limit theorem

- Z = sum of *n* iid RVs X_i
- What is the mean/variance/skew/kurtosis of Z? in terms of these moments of X?

$$Mean[\sum_{i=1}^{n} X_i] = n \cdot Mean[X]$$

$$Variance[\sum_{i=1}^{n} X_i] = n \cdot Variance[X]$$

$$Skew[\sum_{i=1}^{n} X_i] = \frac{1}{\sqrt{n}} \cdot Skew[X]$$

$$Ex.Kurtosis[\sum_{i=1}^{n} X_i] = \frac{1}{n} \cdot Ex.Kurtosis[X]$$

- Mean and variance grow with *n*, but Skew and Kurtosis (and higher moments) go to 0 (what they are for a Normal distribution)
- So as *n* grows, the sum becomes more Gaussian.

Expectation + Normal question

Let's say puppies have the following traits:

wagging speed:	`w` *(Hz)*	`~ Norm ( mu=1, sd=0.25 )`
ear stiffness:	`k` *(log GPa)*	`~ Norm ( mu=-2.5, sd=0.5 )`
eye/head size:	`e` *(log m^2/m^2)*	`~ Norm ( mu=-1.5, sd=1/3 )`

Remarkably, all of these are independent.

We define a cuteness index (λ) for a given puppy as

$$\lambda = 4*w - k + 3*e + 5$$

1) What is the mean λ for all puppies?
2) What is the variance of λ for puppies?
3) What is the probability that a randomly sampled puppy has a cuteness index greater than 10?

Z scores

- What is the probability that our sample mean will have a Z-score > 1.96 or < -1.96?
  (i.e. will be more than 1.96 standard errors away from the population mean?)
- What is the 'critical' absolute Z value such that the Z-score of our sample mean will have an absolute value less than that with probability 68.27%?

Sampling dist. of the sample mean

- We draw N samples from a distribution with mean=100, sd=15 (e.g., N IQ scores). We calculate the mean of those n samples. What is the distribution of the sample mean? (Mean[sample mean]? Variance[sample mean]?)