MATH/STAT 338: Probability

Continuous Random Variables

Anthony Scotina

Probability Density Functions

Continuous vs Discrete

Numerical (Quantitative) Variables

- **Discrete**: Numerical variable that can *only take whole, non-negative numbers* (0, 1, 2, ...)
  - Examples: number of students in STAT 118, number of heads when flipping 3 coins.
- **Continuous**: Numerical variable that can *take an infinite range of numbers* within a (sometimes infinite) interval.
  - Examples: age, temperature, height, weight

We can find $P(X = 2)$ if X is **discrete** (whatever X may be).

- But if Y is **continuous** and can *take an infinite range of numbers*, what is $P(Y = 2)$?

Cumulative Distribution Function

Let Y denote any random variable. The **cumulative distribution function (CDF)** of Y, denoted by $F(Y)$, is such that

$$F(y) = P(Y \leq y), \quad -\infty < y < \infty$$

For $Y \sim Binomial(n = 2, p = 0.5)$...

- $P(Y \leq 0) = P(Y = 0) = \binom{2}{0} 0.5^0 (1 - 0.5)^2 = \boxed{0.25}$
- $P(Y \leq 1) = P(Y = 0) + P(Y = 1) = 0.25 + \binom{2}{1} 0.5^1 (1 - 0.5)^1 = \boxed{0.75}$
- $P(Y \leq 2) = P(Y = 0) + P(Y = 1) + P(Y = 2) = \boxed{1}$

Cumulative Distribution Function

Let Y denote any random variable. The **cumulative distribution function (CDF)** of Y, denoted by $F(Y)$, is such that

$$F(y) = P(Y \leq y), \quad -\infty < y < \infty$$

For $Y \sim Binomial(n = 2, p = 0.5)$...

$$P(Y \leq y) = \begin{cases} 0, & \text{for } y < 0 \\ 0.25, & \text{for } 0 \leq y < 1 \\ 0.75, & \text{for } 1 \leq y < 2 \\ 1, & \text{for } y \geq 2 \end{cases}$$

What does this look like as a graph?

CDF Properties

If $F(y)$ is a *cumulative distribution function*, then:

1. $F(-\infty) = \lim_{y \to -\infty} F(y) = 0$
2. $F(\infty) = \lim_{y \to \infty} F(y) = 1$
3. $F(y)$ is a **nondecreasing** function of y. [If y_1 and y_2 are *any* values that that $y_1 < y_2$, then $F(y_1) \leq F(y_2)$.]

Phone Example

Instead of modeling (for example) the **number** of calls received in an hour, what if we modeled the **lengths** of calls?

- The **number** of calls *has* to be an **integer** (0, 1, 2, ...)
- The **lengths** of calls can be *any* **real number** $c \in [0, \infty)$
  - (Though let's assume the caller gets tired and hangs up after a *max* of 10 mins.)

Suppose we are modeling the **lengths** of calls as a random variable, Y:

> A random variable *Y* with distribution function *F(y)* is said to be **continuous** if *F(y)* is...*continuous*!
>
> - i.e., **not** a *step function*

What does this look like as a graph?

Probability Density Function

Recall that the **probability mass function** gives $P(Y = y)$ for *any* **discrete** random variable Y.

- Because $P(Y = y) = 0$ for any **continuous** random variable, Y, we need something different here...

Let $F(y)$ be the distribution function for a *continuous* random variable Y. Then $f(y)$, given by

$$f(y) = \frac{dF(y)}{dy} = F'(y)$$

wherever the derivative exists, is called the **probability density function (PDF)** for the random variable Y.

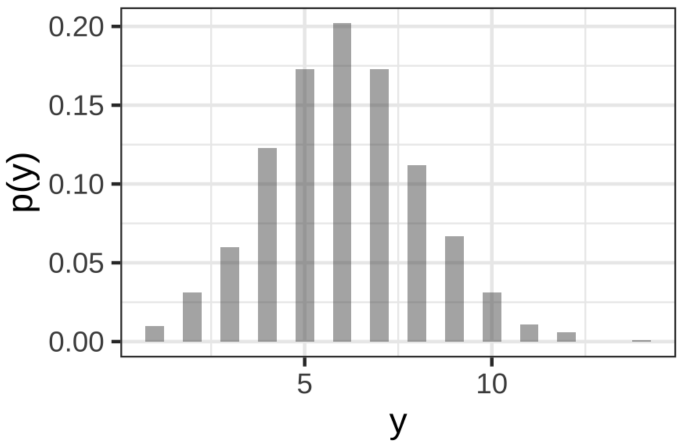
- Therefore, we can find $F(y) = P(Y \leq y)$ for a continuous variable with **integration**:

$$F(y) = \int_{-\infty}^{y} f(t)\, dt$$

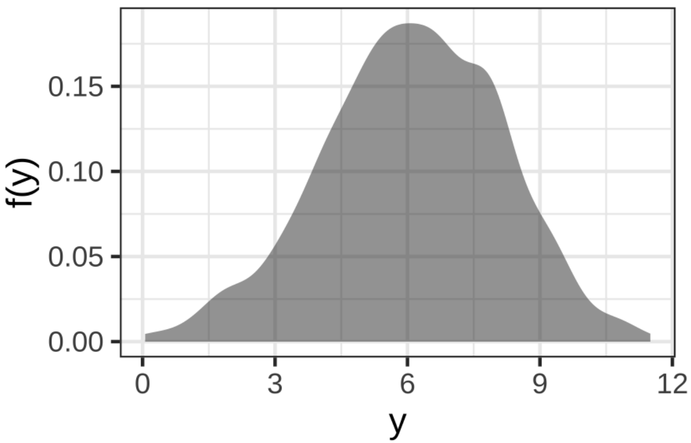
Probability Density Function

Probability density functions, like *probability mass functions*, are *theoretical models for some frequency distribution of a population*:

The **Binomial(20, 0.3)** distribution:



The **Normal(6, 4.2)** distribution:



> *All models are wrong, but some are useful.*
>
> - George Box

PDF Properties

If $f(y)$ is a **density function** for a *continuous* random variable, then:

1. $f(y) \geq 0$ for all y, $-\infty < y < \infty$.
2. $\int_{-\infty}^{\infty} f(y)\, dy = 1$

Compare to PMF properties for *discrete* RVs

For *any* discrete probability distribution, the following must be true:

1. $0 \leq p(y) \leq 1$ for all y.
2. $\sum_y p(y) = 1$, where the summation is over all values of y with nonzero probability.

CAUTION 🚨

🚨The quantity $f(y)$ is **NOT** a probability!!!🚨

- Remember, $P(Y = y) = 0$ for any *continuous* RV.

Then how can we calculate probabilities for continuous RVs?! 🤔

🤗We **integrate** the PDF over the *appropriate range*:

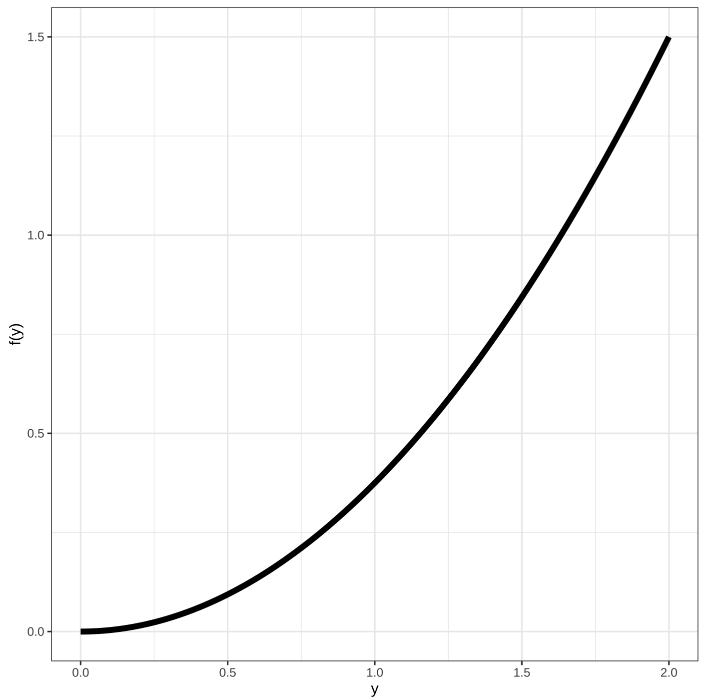
- **Theorem**: If the random variable Y has density function $f(y)$ and CDF $F(y)$, and $a < b$, then the probability that Y falls in the *interval* $[a, b]$ is

$$P(a \leq Y \leq b) = \int_a^b f(y)\, dy = F(b) - F(a)$$

Probabilities with Continuous RVs

Example

Given $f(y) = cy^2, 0 \leq y \leq 2$, and $f(y) = 0$ elsewhere, find the value of c for which $f(y)$ is a valid density function.



Using this valid PDF, find

$$P(1 \leq Y \leq 2)$$

- What about $P(1 < Y < 2)$?

Expected Value

Expected Value

Discrete RVs

Let y be a *discrete random variable* with PMF $p(y)$. Then the **expected value** of Y, denoted as $E(Y)$, is

$$E(Y) = \sum_y yp(y).$$

Continuous RVs

The **expected value** of a *continuous random variable* Y is

$$E(y) = \int_{-\infty}^{\infty} yf(y)\,dy.$$

- **Note**: Sometimes these integrals will be *very difficult* to find, or even *impossible*! In these cases **R** and **simulations** can help us substantially.

Expected Value Theorems

Let Y be a continuous RV with density function $f(y)$.

1. (*Law of the Unconscious Statistician*) Let $g(Y)$ be a real-valued function of Y. Then

$$E[g(Y)] = \int_{\infty}^{\infty} g(y)f(y)\,dy.$$

2. Let c be a constant. Then

$$E(c) = c.$$

3. Let $g(Y)$ be a function of Y, and c be a constant. Then

$$E[cg(Y)] = cE[g(Y)].$$

4. (*Linearity of Expected Value*) Let $g_1(Y), g_2(Y), \ldots, g_k(Y)$ be k functions of Y. Then

$$E[g_1(Y) + g_2(Y) + \cdots + g_k(Y)] = E[g_1(Y)] + E[g_2(Y)] + \cdots + E[g_k(Y)].$$

Example

(WMS 4.21)

If Y has density function

$$f(y) = \begin{cases} (3/2)y^2 + y, & 0 \leq y \leq 1, \\ 0, & \text{elsewhere} \end{cases}$$

find the **mean** and **variance** of Y.

Solution (for $E(Y)$):

$$\begin{aligned} E(Y) &= \int_0^1 y \left(\frac{3}{2} y^2 + y \right) \, dy \\ &= \int_0^1 \frac{3}{2} y^3 + y^2 \, dy \\ &= \frac{3}{8} y^4 + \frac{1}{3} y^3 \Big|_0^1 \\ &= \boxed{0.71} \end{aligned}$$

Practice

Suppose Y is a continuous random variable with density function

$$f(y) = \begin{cases} y/2, & 0 < y < 2, \\ 0, & \text{elsewhere} \end{cases}$$

- Find $E(Y)$.

The **median** of Y is given by the smallest value such that $F(\phi_{0.5}) = P(Y \leq \phi_{0.5}) = 0.5$. Find the **median** of Y.

Example

The $\text{Normal}(\mu, \sigma^2)$ distribution has PDF

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}$$

For the $\text{Normal}(\mu = 100, \sigma^2 = 15^2)$ distribution, can we find $E(Y)$ by hand? 😨

NOPE, BUT R CAN

```
Y = rnorm(n = 10000, mean = 100, sd = 15)
mean(Y)
```

```
## [1] 100.1011
```

- **Spoiler Alert**: It turns out that $E(Y) = \mu$ if $Y \sim \text{Normal}(\mu, \sigma^2)$!

The Uniform Distribution

Uniform PDF

Continuous random variables following the **uniform distribution** on an interval (a, b) are *completely random numbers* between *a* and *b*.

- **Example**: Assuming the subway I take to Simmons picks me up *anytime* between 8am and 8:10am, the **length of time** I'll have to wait follows a uniform distribution.

If $a < b$, a random variable Y is said to have a continuous **uniform distribution** on the interval (a, b) if the density function of Y is

$$f(y) = \frac{1}{b-a}, \qquad a \leq y \leq b$$

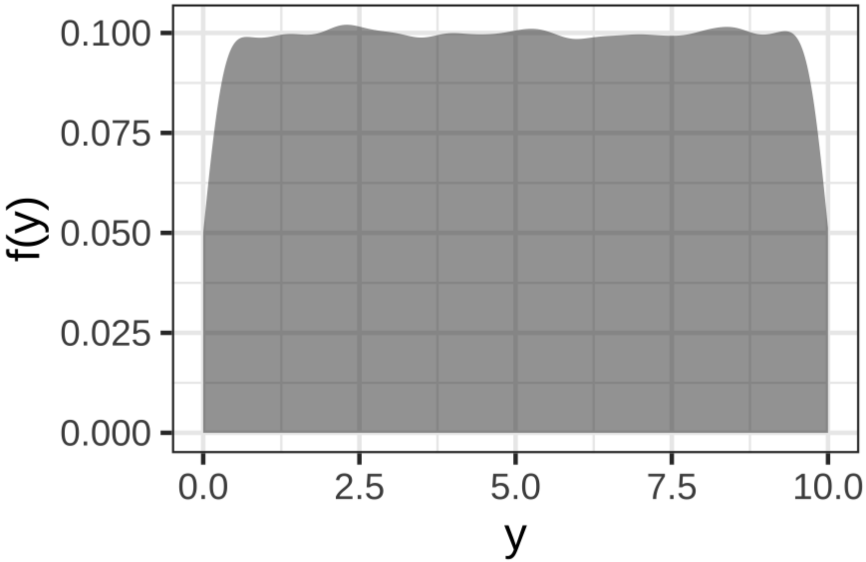
- We typically use $Y \sim \text{Uniform}(a, b)$ as short-hand!

The Uniform Distribution

If $a < b$, a random variable Y is said to have a continuous **uniform distribution** on the interval (a, b) if the density function of Y is

$$f(y) = \frac{1}{b-a}, \qquad a \leq y \leq b$$

The **Uniform(0, 10)** distribution:



A Valid PDF?

Notice that the uniform PDF does *not* depend on values of Y, only on the *length of the interval*:

$$f(y) = \frac{1}{b-a}, \qquad a \leq y \leq b$$

- This means we can prove this is a valid PDF quite easily:

$$\int_{-\infty}^{\infty} f(y)\,dy = \int_a^b \frac{1}{b-a}\,dy = \frac{y}{b-a}\bigg|_a^b = 1$$

Note: **Standard uniform** RVs, $Y \sim \text{Uniform}(0,1)$ fall in the interval $(0,1)$, where

$$f(y) = 1, \qquad 0 \leq y \leq 1$$

Expected Value and Variance

Theorem (Expected value and variance of uniform RVs):

If $a < b$ and Y is a random variable uniformly distributed on the interval (a, b), then

$$\mu = E(Y) = \frac{a+b}{2} \qquad \text{and} \qquad \sigma^2 = Var(Y) = \frac{(b-a)^2}{12}$$

In R: Simulating a **Uniform(10, 20)** distribution

```
Y_sim = runif(n = 1000, min = 10, max = 20)
mean(Y_sim) # simulated E(Y)
```

```
## [1] 14.91654
```

```
(20 + 10)/2 # theoretical E(Y)
```

```
## [1] 15
```

Practice

(WMS 4.51)

The cycle time for trucks hauling concrete to a highway construction site is uniformly distributed over the interval 50 to 70 minutes.

> What is the probability that the cycle time exceeds 65 minutes, if it is known that the cycle time exceeds 55 minutes?
>
> [**Hint**: Think *conditional* probability!]

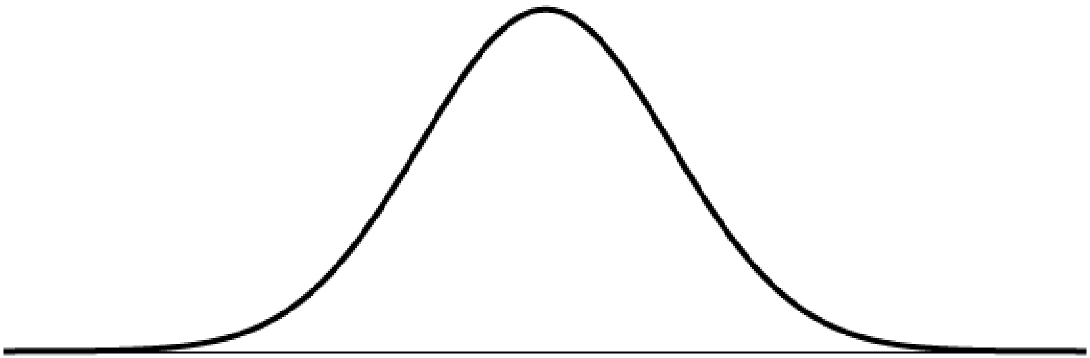
R Simulation:

```
Y_sim = runif(n = 10000, min = 50, max = 70)
mean(Y_sim[Y_sim > 55] > 65)
```

```
## [1] 0.3322264
```

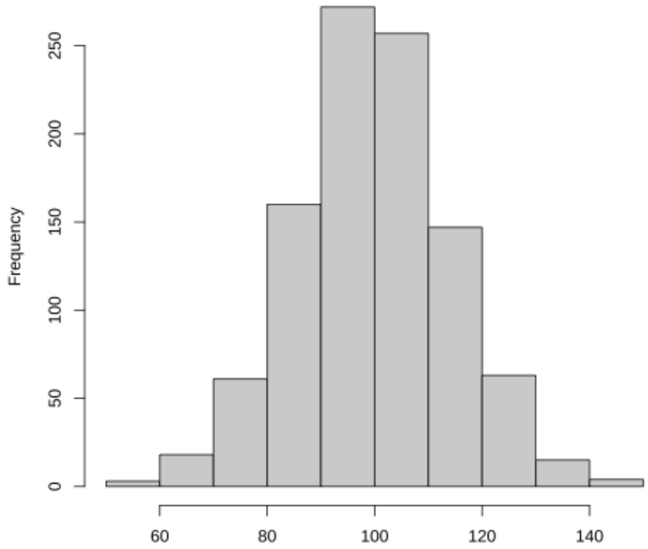
The Normal Distribution

What is it?

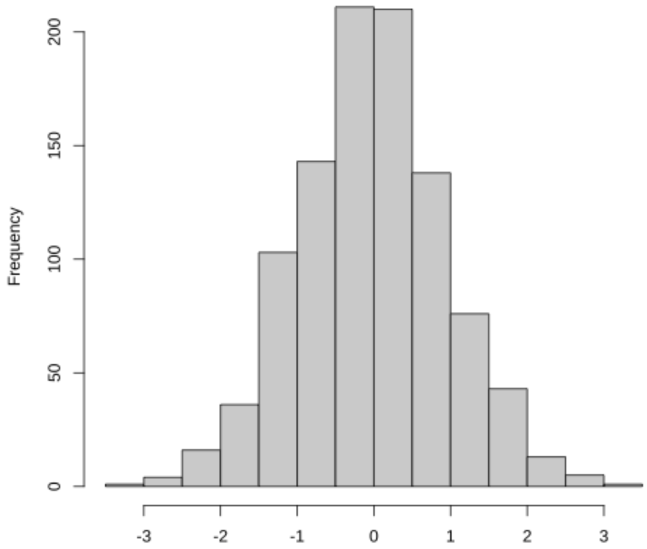


- **unimodal**, **symmetric**
- **bell-shaped**
- Area under the Normal curve **adds up to 1**
- Arises in **many applications**

Different Normal Distributions

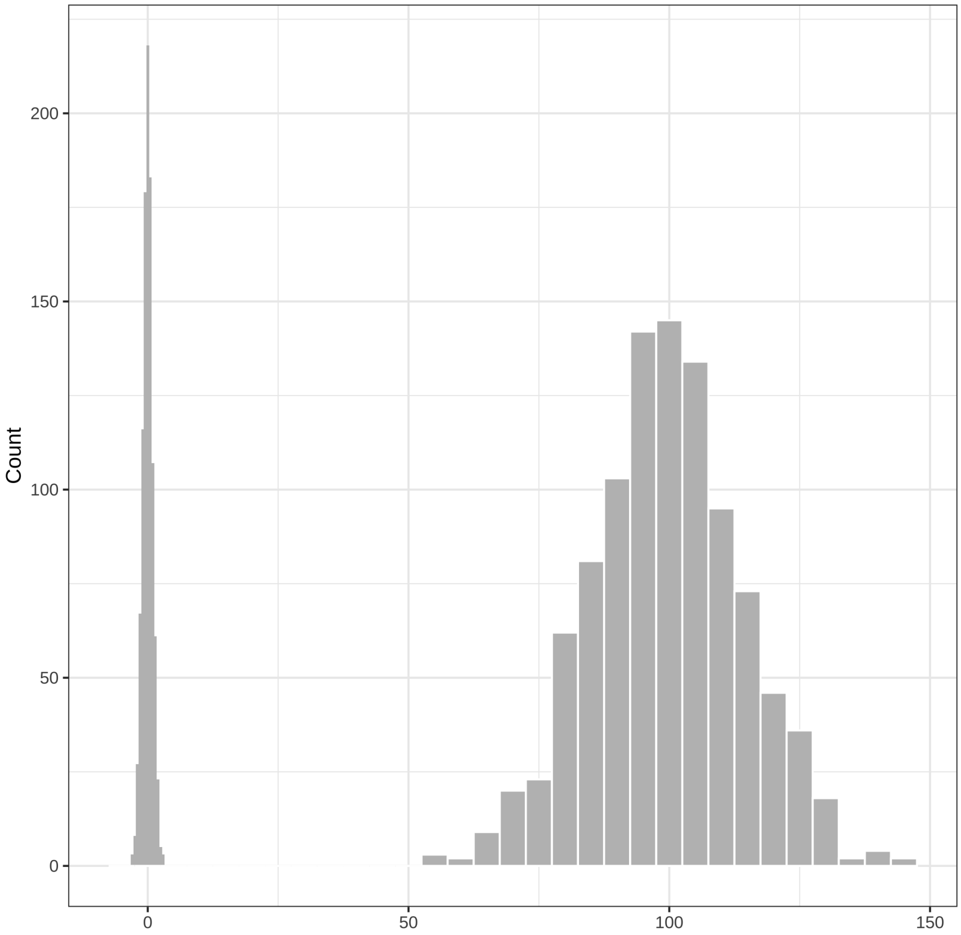


- N(mean = 100, sd = 15)



- N(mean = 0, sd = 1)
- **Standard Normal distribution**

Different Normal Distributions



The Normal PDF

A random variable Y is said to have a **normal** distribution if, for $\sigma > 0$ and $-\infty < \mu < \infty$, the density function of Y is

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-\mu)^2/(2\sigma^2)} \qquad -\infty < y < \infty$$

- We typically use $Y \sim \text{Normal}(\mu, \sigma)$ as short-hand!

Theorem (Expected value and variance of normal RVs):

If Y is normally distributed with parameters μ and σ, then

$$E(Y) = \mu \qquad \text{and} \qquad Var(Y) = \sigma^2.$$

- This means that $SD(Y) = \sigma$.

The Standard Normal PDF

If $Y \sim \text{Normal}(\mu, \sigma)$, we can **shift** and **scale** the normal distribution so that is:

- is centered at 0
- has variance 1

In other words…

$$Z = \frac{Y - \mu}{\sigma} \sim \text{Normal}(0, 1)$$

follows the **standard normal** distribution.

If $Z \sim \text{Normal}(0, 1)$, then

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \qquad -\infty < z < \infty$$

Normal Probabilities

Suppose $Y \sim \mathrm{Normal}(\mu, \sigma)$, and we want to find $P(a \leq Y \leq b)$. This would require us to evaluate

$$\int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-\mu)^2/(2\sigma^2)},$$

which... **we can't do**!!!

Instead, we'll use ~~normal tables~~ **R**. 👍

- `pnorm(y, mu = ..., sd = ...)` computes the **normal CDF**,

$$P(Y \leq y) = \int_{-\infty}^{y} \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-\mu)^2/(2\sigma^2)}$$

Normal Quantiles

Suppose $Y \sim \mathrm{Normal}(\mu, \sigma)$, and we want to find $P(a \leq Y \leq b)$. This would require us to evaluate

$$\int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-\mu)^2/(2\sigma^2)},$$

which... **we can't do**!!!

Instead, we'll use ~~normal tables~~ **R**. 👍

- `qnorm(p, mean = ..., sd = ...)` computes the pth **normal quantile**, which is the value ϕ_p such that

$$P(Y \leq \phi_p) = p, \qquad 0 < p < 1$$

- We can use this to find normal *percentiles*, such as the **median**, where $p = 0.5$.

Practice

(WMS 4.63)

A company that manufactures and bottles apple juice uses a machine that automatically fills 16-ounce bottles. There is some variation, however, in the amounts of liquid dispensed into the bottles that are filled. The amount dispensed, Y, has been observed to be approximately normally distributed with mean 16 ounces and standard deviation 1 ounce.

1. Find the probability that a randomly selected bottle will have more than 17 ounces dispensed into it.

2. Find the *75th percentile* of amounts dispensed; that is, find ϕ_{75} such that $P(Y \leq \phi_{75}) = 0.75$.

Practice

(WMS 4.63)

A company that manufactures and bottles apple juice uses a machine that automatically fills 16-ounce bottles. There is some variation, however, in the amounts of liquid dispensed into the bottles that are filled. The amount dispensed, Y, has been observed to be approximately normally distributed with mean 16 ounces and standard deviation 1 ounce.

1. Find the probability that a randomly selected bottle will have more than 17 ounces dispensed into it.

2. Find the *75th percentile* of amounts dispensed; that is, find ϕ_{75} such that $P(Y \leq \phi_{75}) = 0.75$.

Solution

```r
1 - pnorm(17, mean = 16, sd = 1)
```

```
## [1] 0.1586553
```

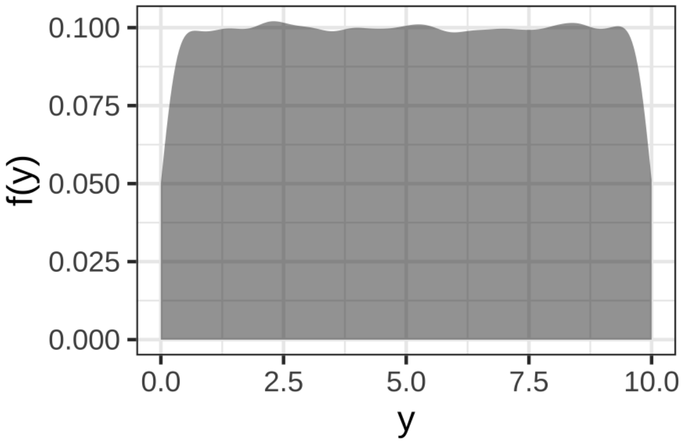
```r
qnorm(0.75, mean = 16, sd = 1)
```

```
## [1] 16.67449
```

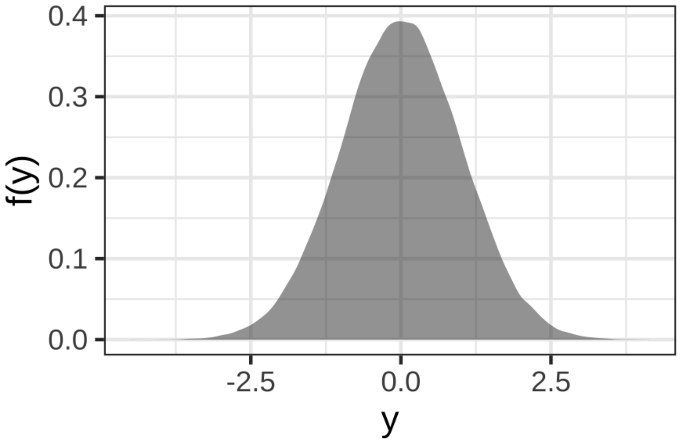
The Gamma Distribution

Non-skewed Distributions

The **Uniform(0, 10)** distribution:

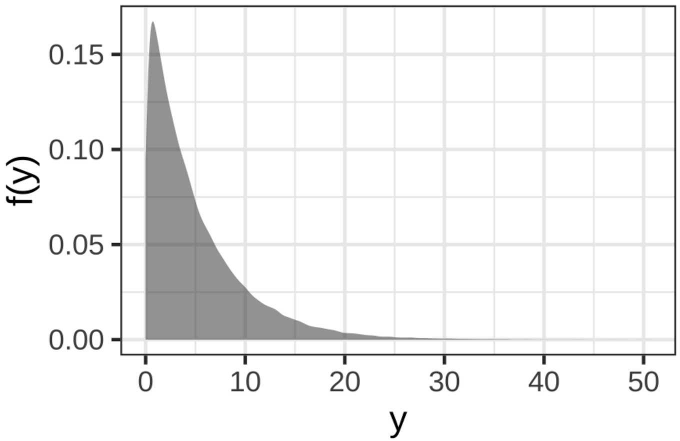


The **Normal(0, 1)** distribution:

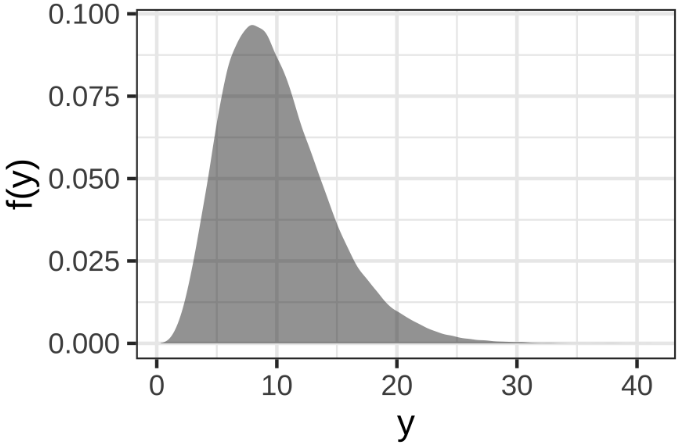


Skewed Distributions

The **Gamma(1, 5)** distribution:



The **Gamma(5, 2)** distribution:



The Gamma PDF

A random variable Y is said to have a **gamma distribution** with *shape* parameter $\alpha > 0$ and *scale* parameter $\beta > 0$ if

$$f(y) = \frac{1}{\beta^\alpha \Gamma(\alpha)} y^{\alpha-1} e^{-y/\beta}, \qquad 0 \leq y < \infty,$$

where

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y}\, dy$$

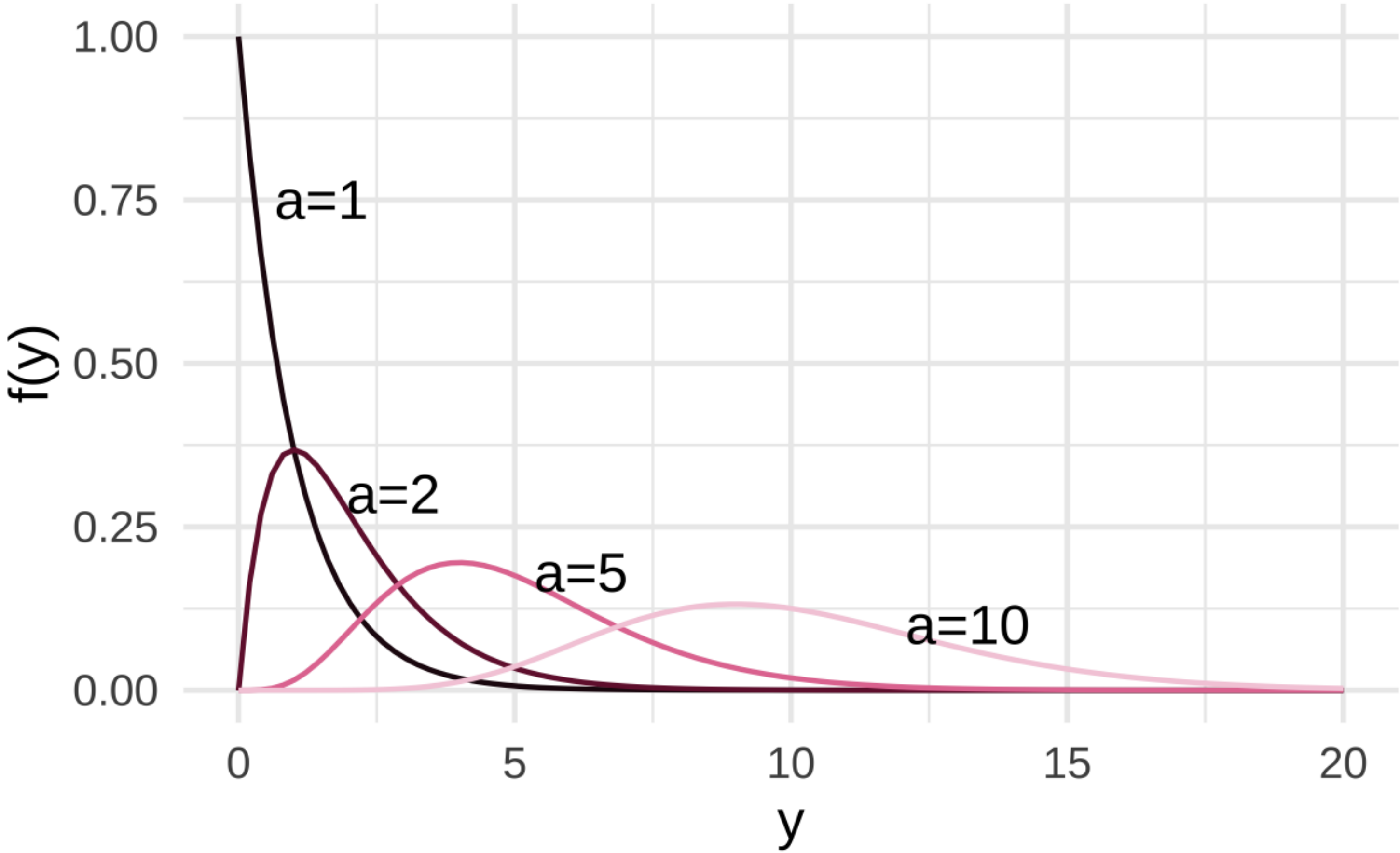
- We typically use $Y \sim \text{Gamma}(\alpha, \beta)$ as short-hand!

The quantity $\Gamma(\alpha)$ is known as the **gamma function**, and it follows that:

- $\Gamma(1) = 1$
- $\Gamma(n) = (n-1)!$, provided that n is an integer
- $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ for any $\alpha > 1$

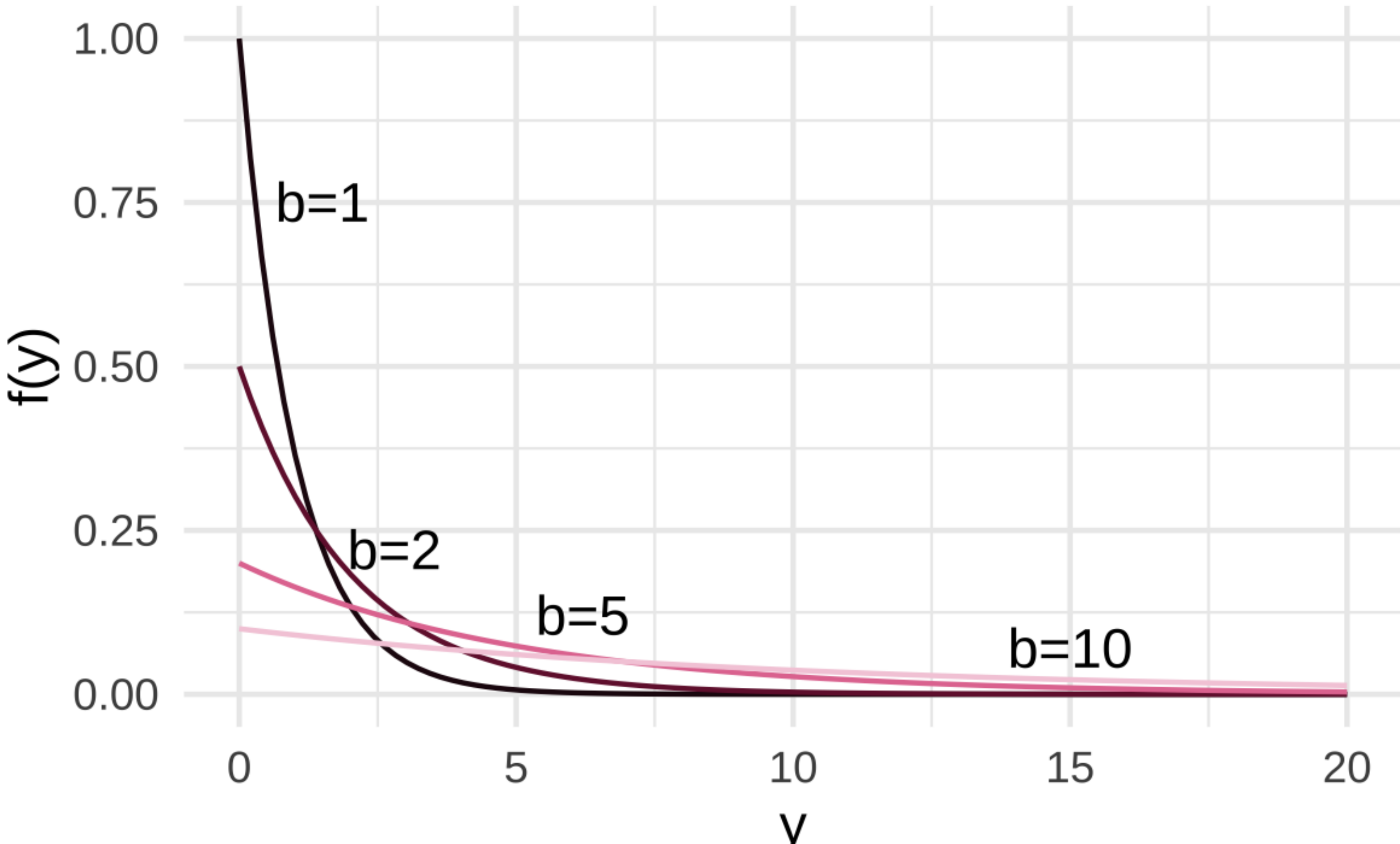
The Shape Parameter, α

- Scale parameter $\beta = 1$



The Scale Parameter, β

- Shape parameter $\alpha = 1$



Expected Value and Variance

Theorem (Expected value and variance of gamma RVs):

If Y has a gamma distribution with parameters α and β, then

$$\mu = E(Y) = \alpha\beta \qquad \text{and} \qquad \sigma^2 = Var(Y) = \alpha\beta^2$$

Expected Value and Variance

Simulations

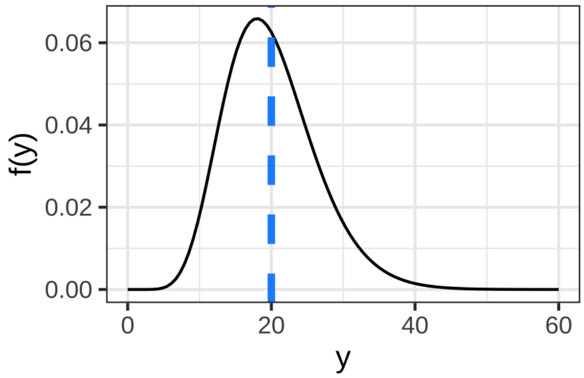
- $Y \sim \text{Gamma}(10, 2)$

```r
Y = rgamma(n = 10000, shape = 10, scale = 2)
mean(Y)
```

```
## [1] 20.04872
```

```r
var(Y)
```

```
## [1] 40.45812
```



Exponential Distribution

The **exponential distribution** is a *special case* of the gamma distribution where $\alpha = 1$.

- That is, $Y \sim \text{Gamma}(1, \beta)$ is the same as $Y \sim \text{Exponential}(\beta)$.

A random variable Y is said to have an **exponential distribution** with parameter $\beta > 0$ if the density function of Y is

$$f(y) = \frac{1}{\beta} e^{-y/\beta}, \qquad 0 \leq y < \infty$$

This is similar to the **geometric distribution**!

- **Geometric distribution**: Models the the **number** of the trial on which the *first success* occurs
- **Exponential distribution**: Models the first success in **continuous time**
  - E.g., the *waiting time* until the first "success"

Survival Analysis Detour

The exponential distribution is common in **survival analysis**, which is the analysis of *time-to-event* data.

Suppose we want to model the *time-to-event*, T, in a study of a treatment for type 2 diabetes (an example of the "event" could be myocardial infarction).

- The **survival function**, $S(t)$, gives the probability that a study participant can "survive" beyond time t:

$$S(t) = 1 - F(t) = P(T > t) = \int_t^\infty f(x)\,dx$$

- Let's find $S(t)$ if $T \sim \mathrm{Exponential}(\beta)$.

Survival Analysis Detour

The exponential distribution is common in **survival analysis**, which is the analysis of *time-to-event* data.

Suppose we want to model the *time-to-event*, T, in a study of a treatment for type 2 diabetes (an example of the "event" could be myocardial infarction).

- The **hazard function**, $h(t)$, provides a *conditional density* of an event, given that the event *has not yet occurred* prior to time t:

$$h(t) = \frac{f(t)}{S(t)}$$

- **Example**: Given that I haven't experienced symptoms yet, what are my chances of experiencing them in the next year?

Practice: Find $h(t)$ for $T \sim \text{Exponential}(\beta)$.

- **Note**: The hazard function is **constant** for the exponential distribution.

The Memoryless Property

The exponential distribution has the **memoryless property**.

> Even if you've waited for hours or days without success, the success isn't any more likely to arrive soon. In fact, you might as wekk have just started waiting 10 seconds ago.
>
> - Blitzstein and Hwang, 2019

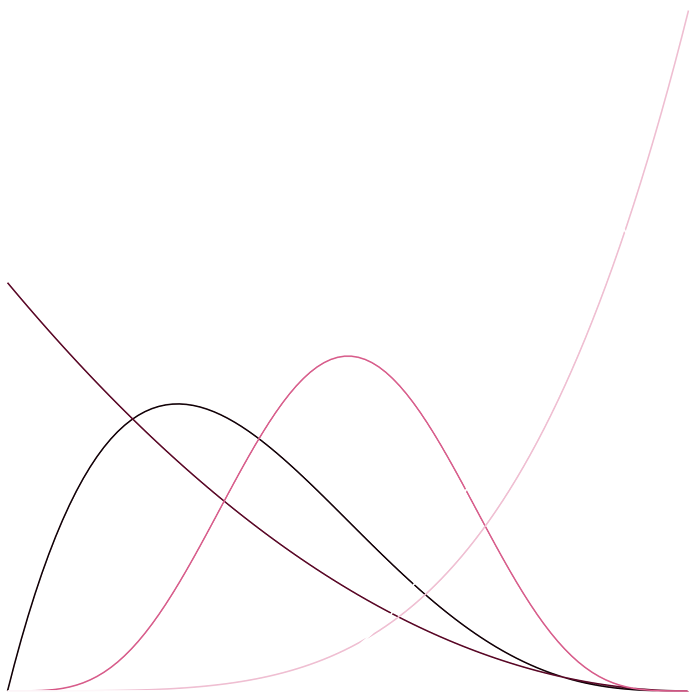
Definition: A continuous distribution has the **memoryless property** if

$$P(Y > a + b \mid Y > a) = P(Y > b).$$

- If a distribution has the memoryless property, then...

$$E(Y \mid Y > a) = a + E(Y)$$

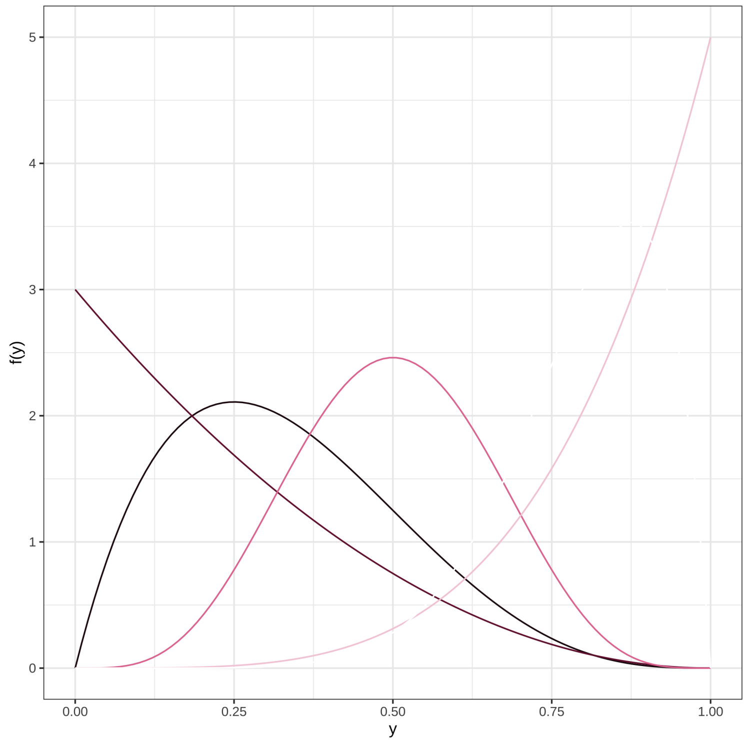
The Beta Distribution



Models for Proportions

The **beta distribution** is used to model *proportions* - variables that can take *only* values between **0** and **1**.

- It is also a generalization of the **Uniform(0, 1)** distribution, and allows for a non-constant PDF over the interval $(0, 1)$.



The Beta PDF

A random variable Y is said to have a **beta distribution** with parameters $\alpha > 0$ and $\beta > 0$ if the density function of Y is

$$f(y) = \frac{1}{B(\alpha, \beta)} y^{\alpha-1}(1-y)^{\beta-1}, \qquad 0 \le y \le 1,$$

where

$$B(\alpha, \beta) = \int_0^1 y^{\alpha-1}(1-y)^{\beta-1}\, dy = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

- We typically use $Y \sim \mathrm{Beta}(\alpha, \beta)$ as short-hand!

Beta Distributions

Beta(0.5, 0.5)

Beta Distributions

Beta(1, 1) = **Uniform(0, 1)**

Beta Distributions

Beta(5, 5)

Beta Distributions

Beta(8, 2)

Beta Distributions

Beta(2, 8)

Expected Value and Variance

Theorem (Expected value and variance of beta RVs):

If Y is a beta-distributed random variable with parameters $\alpha > 0$ and $\beta > 0$, then

$$\mu = E(Y) = \frac{\alpha}{\alpha + \beta}$$

and

$$\sigma^2 = Var(Y) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

Practice

(WMS 4.129)

During an eight-hour shift, the proportion of time Y that a sheet-metal stamping machine is down for maintenance or repairs has a beta distribution with $\alpha = 1$ and $\beta = 2$. That is,

$$f(y) = 2(1 - y), \quad 0 \leq y \leq 1.$$

The cost (in hundreds of dollars) of this downtime, due to lost production and cost of maintenance and repair, is given by $C = 10 + 20Y + 4Y^2$.

> Find the mean (by hand) and the variance (using a simulation) of C.

Simulation

```r
Y = rbeta(n = 10000, shape1 = 1, shape2 = 2)
C = 10 + 20*Y + 4*Y^2

mean(C)
```

```
## [1] 17.35881
```

```r
var(C)
```

The Beta-Binomial Conjugacy

The story

Suppose we are tossing a **biased coin**, but we don't know *how* biased it is.

- In other words, it lands Heads with some *unknown* probability, p.
- We could *infer* the value of p by tossing the coin a bunch of times. But that isn't **fun**...

In **Bayesian inference**, we treat all unknown quantities, including p, as random variables!

- Thus, p would have its own *probability distribution*, called a **prior distribution**.
    - This distribution reflects our *uncertainty* about the true value of p before actually observing a bunch of coin tosses.
- After tossing the coin bunch of times and we have a better idea of the value of p (i.e., we collect **data**), we update the prior distribution with Bayes' Rule.
    - This yields the **posterior distribution**.

The Beta-Binomial Conjugacy

Let's assign the $Beta(\alpha, \beta)$ distribution to the *unknown* p.

- That is, $p \sim \text{Beta}(\alpha, \beta)$.

Let Y be the number of Heads in n tosses of the coin.

- This **depends on** p, so we would write:

$$Y \mid p \sim \text{Binomial}(n, p)$$

To update the **prior distribution** of p based on our "data" (Y), we use Bayes' Rule:

$$f(p \mid Y = y) = \frac{P(Y = y \mid p) f(p)}{P(Y = y)}$$

We'll look at **conditional distributions** later, but it turns out that

$$p \mid Y = y \sim \text{Beta}(\alpha + y, \beta + n - y)$$

- That is, if our *prior beliefs* follow the beta distribution, and our **data** follow the binomial distribution, then our *posterior distribution* **still** follows the beta!

Moment-Generating Functions

Moments

Moments provide us a set of *additional* measures beyond μ and σ^2 that (usually) *uniquely determine* a probability distribution.

- The kth **moment** of a random variable Y is defined to be $E(Y^k)$ and is denoted by $\mu_k^{'}$.
- The kth **central moment** of a random variable Y is defined to be $E[(Y-\mu)^k]$ and is denoted by μ_k.

We've seen some **moments** already!

- $E(Y) = \mu_1^{'} = \mu$ is the *first moment.*
- $E(Y^2)$ is the *second moment.*
  - This is used to find $Var(Y) = E[(Y-\mu)^2] = \mu_2$, the *second central moment.*
- The **skewness** of a random variable $\boxed{\&}$ is the third *standardized moment,*

Moment-Generating Functions

For a continuous random variable Y, the **moment-generating function (MGF)**, $m(t)$, is defined to be

$$m(t) = E(e^{tY})$$

.

- We can evaluate this for any **continuous distribution** with an integral:

$$m(t) = E(e^{tY}) = \int_{-\infty}^{\infty} e^{ty} f(y)\, dy$$

Theorem:

If $m(t)$ exists, then for any positive integer k,

$$\left.\frac{d^k m(t)}{dt^k}\right]_{t=0} = m^{(k)}(0) = \mu_k^{'}$$

- In other words, find the kth derivative of $m(t)$, plug in 0 for t, and you're left with $\mu_k^{'}$.

Gamma MGF

The **Gamma** PDF with *shape* parameter $\alpha > 0$ and *scale* parameter $\beta > 0$ if

$$f(y) = \frac{1}{\beta^\alpha \Gamma(\alpha)} y^{\alpha-1} e^{-y/\beta}, \qquad 0 \le y < \infty$$

Let's derive the **Gamma** MGF:

$$m(t) = E(e^{tY}) = \int_0^\infty e^{ty} \left[\frac{y^{\alpha-1} e^{-y/\beta}}{\beta^\alpha \Gamma(\alpha)} \right] dy = \frac{1}{(1-\beta t)^\alpha}$$

Normal MGF

The **Normal** PDF with mean μ and standard deviation σ is given by:

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-\mu)^2/(2\sigma^2)} \qquad -\infty < y < \infty$$

The *moment-generating function* for this distribution is:

$$m(t) = e^{\mu t + t^2\sigma^2/2}$$

- Let's use this to find $E(X^3)$.