CSCI 3022-002 Intro to Data Science
Expectation

Example:

Suppose a light bulb's lifetime is exponentially distributed with parameter λ.

One (often) appealing property of the exponential is its *memoryless property*. In particular, consider the knowledge gained by knowing that the "event" has not yet occurred by time t_0. What is $P\left(X > (t_0 + t) | X > t_0\right)$?

'Memoryless'

For $X \sim exp(\lambda)$, what is $P\left(X > (t_0 + t)|X > t_0\right)$?

'Memoryless'

For $X \sim exp(\lambda)$, what is $P\left(X > (t_0 + t) | X > t_0\right)$?

$$P\left(X > (t_0 + t) | X > t_0\right) = \frac{P\left(X > (t_0 + t) \text{ and } X > t_0\right)}{P\left(X > t_0\right)}$$

then use that $F(x) = 1 - e^{-\lambda x}$:

$$= \frac{1 - (1 - e^{\lambda(t_0+t)})}{1 - (1 - e^{\lambda t_0})}$$

$$= \frac{e^{\lambda(t_0+t)}}{e^{\lambda t_0}} = e^{\lambda t} = P(X > t)$$

Or we've gained no knowledge about future burnout time of the light based on the past t_0!

Announcements and Reminders

- Practicum due next Monday!

EV Recap

1. **Expected Value:** The average value for X coming from a distribution (not a sample!). Denoted $E[X]$ or μ or μ_X.
   Discrete: $\sum_{x \in \Omega} x f(x)$; Continuous: $\int_{x \in \Omega} x \cdot f(x)\, dx$
2. Expected value of a function $g(X)$ of X is:
   $\sum_{x \in \Omega} g(x) f(x)$; $\int_{x \in \Omega} g(x) \cdot f(x)\, dx$
3. $Y = g(X)$ is a *change of variables*.
4. Expectation is **linear:** $E[aX + b] = aE[X] + b$

Plinko... is random?

Let X be the random variable describing the result in each round of Plinko with n rows and probability p of moving to the right off of each peg. (Ignoring the edges for now.) What is the *Variance* of X?

Recall: Sample *Variance* is $\dfrac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n-1}$

Another way: sample variance is $\underbrace{\dfrac{1}{n-1}\sum_{i=1}^{n}}_{\text{averaged out}} \underbrace{\left(X_i - \bar{X}\right)^2}_{\text{squared deviations}}$

Population variance is this idea expressed as an expectation:

$$Var[X] = E[\underbrace{(X - E[X])^2}_{\text{squared deviations}}] = E[(X - \mu_X)^2]$$

Variance of a Random Variable

Definition: *Variance:*
For a discrete random variable X with pdf $f(x)$, the *variance* of X is denoted as ______________ and is calculated as:

1. Continuous:

2. Discrete:

The standard deviation (SD) of X is:

Variance of a Random Variable

Definition: *Variance:*
For a discrete random variable X with pdf $f(x)$, the *variance* of X is denoted as $\underline{Var[X] = \sigma^2}$ and is calculated as:

$$Var[X] = E[(X - E[X])^2]$$

1. Continuous:

$$Var[X] = \int_{x\in\Omega} (x - \mu_x)^2 \cdot f(x)\, dx$$

2. Discrete:

$$Var[X] = \sum_{x\in\Omega} (x - \mu_x)^2 f(x)$$

The standard deviation (SD) of X is: $\sigma = \sqrt{\sigma^2}$

Variance Calculated

We want more Plinko! Let's find the variance of a Bernoulli so we can build on it.

Recall: The pmf of the Bernoulli is given by

$$f(x) = \begin{cases} p & x = 1 \\ 1 - p & x = 0 \end{cases}$$

and we know that $E[X] = p$.

Variance Calculated

We want more Plinko! Let's find the variance of a Bernoulli so we can build on it.

Recall: The pmf of the Bernoulli is given by

$$f(x) = \begin{cases} p & x = 1 \\ 1-p & x = 0 \end{cases}$$

and we know that $E[X] = p$. We now must sum over both outcomes' deviations from the mean while multiplying by those probabilities

$$E[(X - E[X])^2] = \sum_{x \in \{0,1\}} (x-p)^2 f(x) = \sum_{x \in \{0,1\}} (x-p)^2 P(X = x)$$

$$= (0-p)^2 \cdot P(X=0) + (1-p)^2 \cdot P(X=1) = (0-p)^2 \cdot (1-p) + (1-p)^2 \cdot p$$

$$= (p)(1-p)(p+1-p) = \boxed{p(1-p)}$$

Let's play Plinko!!

Let X be the random variable describing the result in each round of Plinko with n rows and probability p of moving to the right off of each peg. (Ignoring the edges for now.) What is the *variance* of X?

Need to know: if two random variables are **independent**,

$$\boxed{Var[X+Y] = Var[X] + Var[Y]}$$

Let's play Plinko!!

Let X be the random variable describing the result in each round of Plinko with n rows and probability p of moving to the right off of each peg. (Ignoring the edges for now.) What is the *variance* of X?

Need to know: if two random variables are **independent**,
$\boxed{Var[X+Y] = Var[X] + Var[Y]}$

So for Plinko, where $X = Y_1 + Y_2 + \cdots + Y_i$ but the $Y_i's$ are all independent,

$$Var[X] = Var[\sum Y_i] \underbrace{=}_{\text{indep}} \sum Var[Y] \underbrace{=}_{\text{ident}} nVar[Y_i] = \boxed{np(1-p)}$$

Sanity Check! Should variance be smaller if $p \approx 1$ or $p \approx 0$?

Let's talk Variance

For a random variable X and constants a and b, if we define $Y = aX + b$...
$E[Y] = aE[X] + b$ because Expectation $E[\cdot]$ is **linear**. Is $Var[\cdot]$?

1. What is $Var[X + b]$?

2. What is $Var[aX]$?

Let's talk Variance

For a random variable X and constants a and b, if we define $Y = aX + b$...
$E[Y] = aE[X] + b$ because Expectation $E[\cdot]$ is **linear**. Is $Var[\cdot]$?

1. What is $Var[X + b]$?
   Intuition: moving X doesn't change its spread!

2. What is $Var[aX]$?
   Intuition: multiplying X should change its spread!

Non-linear Variance

For a random variable X and constants a and b, if we define $Y = aX + b$...

What is $Var[aX + b]$?

Non-linear Variance

For a random variable X and constants a and b, if we define $Y = aX + b$...

What is $Var[aX + b]$?

$$\begin{aligned}
Var[aX + b] &= \sum_{x \in \Omega} (aX + b - E[aX + b])^2 f(x) \\
&= \sum_{x \in \Omega} (aX + b - aE[X] - b)^2 f(x) \\
&= \sum_{x \in \Omega} (aX - aE[X])^2 f(x) \\
&= \sum_{x \in \Omega} a^2 (X - E[X])^2 f(x) \\
&= a^2 \sum_{x \in \Omega} (X - E[X])^2 f(x) \\
&= \boxed{a^2 Var[X]}
\end{aligned}$$

Really non-linear Variance

What if we want to know what happens to two events that *aren't* independent? For example, what's the variance of $Z = X + Y$?

Really non-linear Variance

What if we want to know what happens to two events that *aren't* independent? For example, what's the variance of $Z = X + Y$?

$$Var[X+Y] = \sum_{x \in \Omega} (X + Y - E[X+Y])^2 f(x)$$

$$Var[X+Y] = \sum_{x \in \Omega} (X + Y - E[X] - E[Y])^2 f(x)$$

Really non-linear Variance

What if we want to know what happens to two events that *aren't* independent? For example, what's the variance of $Z = X + Y$?

$$Var[X+Y] = \sum_{x \in \Omega} (X + Y - E[X+Y])^2 f(x)$$

$$Var[X+Y] = \sum_{x \in \Omega} (X + Y - E[X] - E[Y])^2 f(x)$$

If we expand this out, we have to deal with a bunch of $XY, XE[Y]$, etc. terms. It matters if X and Y move *together*. It helps to define this concept. What does it mean for X and Y to move *together*?
Example: what if $Y = -X$? Then the variance of Z is zero!

Covariance

When two random variables X and Y are not independent, it is frequently of interest to assess how strongly they are related to one another.
Definition: *Covariance:*
The covariance between two rv's X and Y is defined as:

$$E[\ \underbrace{(X - \mu_X)}_{\text{X versus its mean}}\ \overbrace{(Y - \mu_Y)}^{\text{Y versus its mean}}\ ]$$

If both variables tend to deviate in the same direction (both go above their means or below their means at the same time), then the covariance will be positive.
If the opposite is true, the covariance will be negative.
If X and Y are not strongly related, the covariance will be near 0.

Correlation

Definition: *Correlation*
The *correlation* coefficient of X and Y, denoted by __________ or just _, is the *unitless* measure of covariance defined by:

It represents a "scaled" covariance: correlation ranges between -1 and 1.

Correlation

Definition: *Correlation*

The *correlation* coefficient of X and Y, denoted by $\underline{Cov[X,Y]}$ or just $\underline{\rho}$, is the *unitless* measure of covariance defined by:

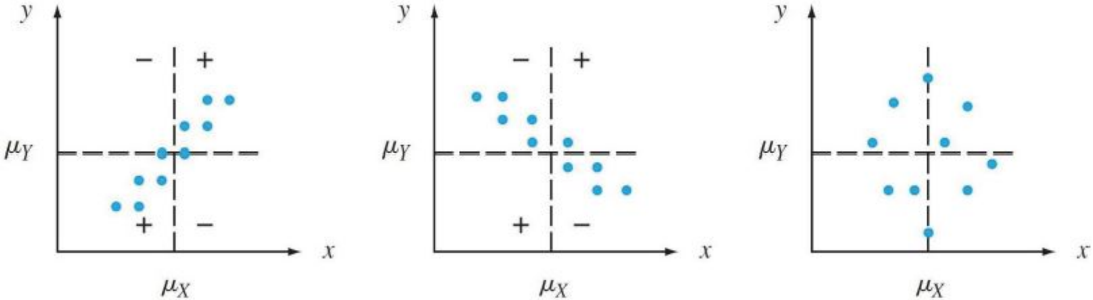
$$\rho = \frac{Cov[X,Y]}{\sigma_X \sigma_Y}$$

It represents a "scaled" covariance: correlation ranges between -1 and 1.

Covariance Pictured

The covariance depends on both the set of possible pairs and the probabilities of those pairs.

Below are examples of 3 types of “co-varying”:



Interpreting Correlation

If X and Y are independent, then _____, but _____ does not imply independence.

The correlation coefficient is a measure of the *linear relationship* between X and Y, and only when the two variables are perfectly related in a *linear* manner will be as positive or negative as it can be.

Two variables could be uncorrelated yet highly dependent because there is a strong nonlinear relationship, so be careful not to conclude too much from low correlation.

We return to covariance in a few weeks...

Interpreting Correlation

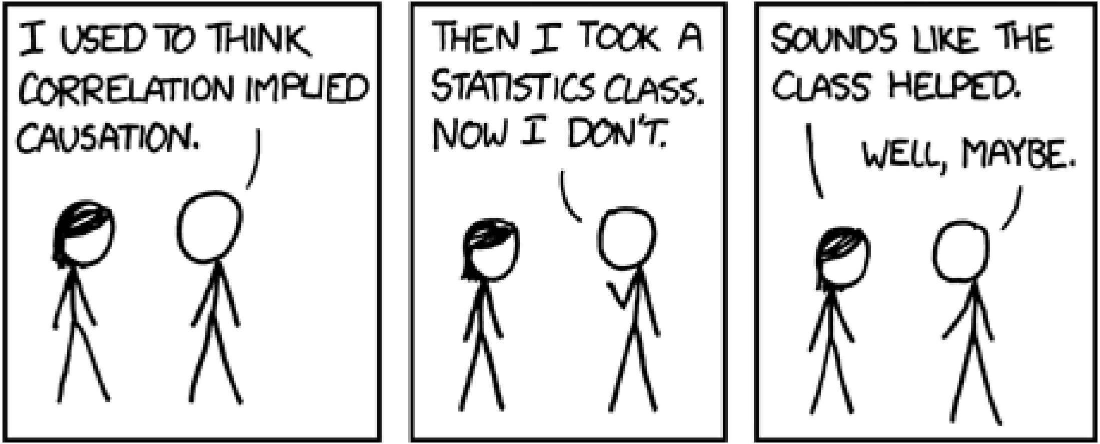
If X and Y are independent, then $\underline{\rho = 0}$, but $\underline{\rho = 0}$ does not imply independence.

The correlation coefficient is a measure of the *linear relationship* between X and Y, and only when the two variables are perfectly related in a *linear* manner will be as positive or negative as it can be.

Two variables could be uncorrelated yet highly dependent because there is a strong nonlinear relationship, so be careful not to conclude too much from low correlation.

We return to covariance in a few weeks...

Interpreting Correlation



Riddles

The Riddler is a column that posts logic, math, and probability puzzles each week. Many can be attacked by simulation and discrete probability: things we can solve!

(`https://fivethirtyeight.com/features/how-many-phones-do-you-need-to-win-hq-trivia/`)

I have a matching game app for my 4-year-old daughter. There are 10 different pairs of cards, each pair depicting the same animal. That makes 20 cards total, all arrayed face down. The goal is to match all the pairs. When you flip two cards up, if they match, they stay up, decreasing the number of unmatched cards and rewarding you with the corresponding animal sound. If they don't match, they both flip back down. (Essentially like Concentration.) However, my 1-year-old son also likes to play the game, exclusively for its animal sounds. He has no ability to match cards intentionally — it's all random.

If he flips a pair of cards every second and it takes another second for them to either flip back over or to make the "matching" sound, how long should my daughter expect to have to wait before he finishes the game and it's her turn again?

Riddler Soln

If he flips a pair of (20) cards every second and it takes another second for them to either flip back over or to make the "matching" sound, how long should my daughter expect to have to wait before he finishes the game and it's her turn again?

1. We are computing $E[time]$ where $time = tries \cdot 2$. Let's compute $E[tries]$.
2. With all 20 cards at random, he picks a card. What's the probability the other card is its match?
3. What about once there are only 18 cards?
4. What random variables are these? What's the expected value of that RV?
5. The total time is the expected sum of these RVs, but that's also the sum of their expectations! Linearity!

Riddler Soln

If he flips a pair of (20) cards every second and it takes another second for them to either flip back over or to make the "matching" sound, how long should my daughter expect to have to wait before he finishes the game and it's her turn again?

1. We are computing $E[time]$ where $time = tries \cdot 2$. Let's compute $E[tries]$.
2. With all 20 cards at random, he picks a card. What's the probability the other card is its match? 1/19
3. What about once there are only 18 cards? 1/17
4. What random variables are these? What's the expected value of that RV? These are **geometric**. The mean number of trials is $1/p$
5. The total time is the expected sum of these RVs, but that's also the sum of their expectations! Linearity!

In practice, we often just look up the formulas for the pdfs, means, and variances of whatever model we choose to use.

Table of Common Distributions

taken from *Statistical Inference* by Casella and Berger

Discrete Distrbutions

distribution	pmf	mean	variance	mgf/moment
Bernoulli(p)	$p^x(1-p)^{1-x};\ x=0,1;\ p\in(0,1)$	p	$p(1-p)$	$(1-p)+pe^t$
Beta-binomial(n,α,β)	$\binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(x+\alpha)\Gamma(n-x+\beta)}{\Gamma(\alpha+\beta+n)}$	$\frac{n\alpha}{\alpha+\beta}$	$\frac{n\alpha\beta}{(\alpha+\beta)^2}$	
Notes: If $X\|P$ is binomial (n,P) and P is beta(α,β), then X is beta-binomial(n,α,β).				
Binomial(n,p)	$\binom{n}{x}p^x(1-p)^{n-x};\ x=1,\ldots,n$	np	$np(1-p)$	$[(1-p)+pe^t]^n$
Discrete Uniform(N)	$\frac{1}{N};\ x=1,\ldots,N$	$\frac{N+1}{2}$	$\frac{(N+1)(N-1)}{12}$	$\frac{1}{N}\sum_{i=1}^{N}e^{it}$
Geometric(p)	$p(1-p)^{x-1};\ p\in(0,1)$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pe^t}{1-(1-p)e^t}$
Note: $Y=X-1$ is negative binomial($1,p$). The distribution is *memoryless*: $P(X>s\|X>t)=P(X>s-t)$.				
Hypergeometric(N,M,K)	$\frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}};\ x=1,\ldots,K$	$\frac{KM}{N}$	$\frac{KM}{N}\frac{(N-M)(N-k)}{N(N-1)}$	?
	$M-(N-K)\le x\le M;\ N,M,K>0$			
Negative Binomial(r,p)	$\binom{r+x-1}{x}p^r(1-p)^x;\ p\in(0,1)$	$\frac{r(1-p)}{p}$	$\frac{r(1-p)}{p^2}$	$\left(\frac{p}{1-(1-p)e^t}\right)^r$
	$\binom{y-1}{r-1}p^r(1-p)^{y-r};\ Y=X+r$			
Poisson(λ)	$\frac{e^{-\lambda}\lambda^x}{x!};\ \lambda\ge 0$	λ	λ	$e^{\lambda(e^t-1)}$
Notes: If Y is gamma(α,β), X is Poisson($\frac{x}{\beta}$), and α is an integer, then $P(X\ge\alpha)=P(Y\le y)$.				

Continuous Distributions

distribution	pdf	mean	variance	mgf/moment
Beta(α, β)	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1};\ x \in (0,1),\ \alpha, \beta > 0$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	$1 + \sum_{k=1}^{\infty} \left(\prod_{r=0}^{k-1} \frac{\alpha+r}{\alpha+\beta+r} \right) \frac{t^k}{k!}$
Cauchy(θ, σ)	$\frac{1}{\pi\sigma} \frac{1}{1+(\frac{x-\theta}{\sigma})^2};\ \sigma > 0$	does not exist	does not exist	does not exist
Notes: Special case of Students's t with 1 degree of freedom. Also, if X, Y are iid $N(0,1)$, $\frac{X}{Y}$ is Cauchy				
χ_p^2	$\frac{1}{\Gamma(\frac{p}{2}) 2^{\frac{p}{2}}} x^{\frac{p}{2}-1} e^{-\frac{x}{2}};\ x > 0,\ p \in N$	p	$2p$	$\left(\frac{1}{1-2t}\right)^{\frac{p}{2}},\ t < \frac{1}{2}$
Notes: Gamma$(\frac{p}{2}, 2)$.				
Double Exponential(μ, σ)	$\frac{1}{2\sigma} e^{-\frac{\lvert x-\mu \rvert}{\sigma}};\ \sigma > 0$	μ	$2\sigma^2$	$\frac{e^{\mu t}}{1-(\sigma t)^2}$
Exponential(θ)	$\frac{1}{\theta} e^{-\frac{x}{\theta}};\ x \geq 0,\ \theta > 0$	θ	θ^2	$\frac{1}{1-\theta t},\ t < \frac{1}{\theta}$
Notes: Gamma$(1, \theta)$. Memoryless. $Y = X^{\frac{1}{\gamma}}$ is *Weibull*. $Y = \sqrt{\frac{2X}{\beta}}$ is *Rayleigh*. $Y = \alpha - \gamma \log \frac{X}{\beta}$ is *Gumbel*.				
F_{ν_1, ν_2}	$\frac{\Gamma(\frac{\nu_1+\nu_2}{2})}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})} \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} \frac{x^{\frac{\nu_1-2}{2}}}{\left(1+(\frac{\nu_1}{\nu_2})x\right)^{\frac{\nu_1+\nu_2}{2}}};\ x > 0$	$\frac{\nu_2}{\nu_2-2},\ \nu_2 > 2$	$2(\frac{\nu_2}{\nu_2-2})^2 \frac{\nu_1+\nu_2-2}{\nu_1(\nu_2-4)},\ \nu_2 > 4$	$EX^n = \frac{\Gamma(\frac{\nu_1+2n}{2})\Gamma(\frac{\nu_2-2n}{2})}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})} \left(\frac{\nu_2}{\nu_1}\right)^n,\ n < \frac{\nu_2}{2}$
Notes: $F_{\nu_1,\nu_2} = \frac{\chi_{\nu_1}^2/\nu_1}{\chi_{\nu_2}^2/\nu_2}$, where the χ^2s are independent. $F_{1,\nu} = t_\nu^2$.				
Gamma(α, β)	$\frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}};\ x > 0,\ \alpha, \beta > 0$	$\alpha\beta$	$\alpha\beta^2$	$\left(\frac{1}{1-\beta t}\right)^{\alpha},\ t < \frac{1}{\beta}$
Notes: Some special cases are exponential $(\alpha = 1)$ and χ^2 $(\alpha = \frac{p}{2}, \beta = 2)$. If $\alpha = \frac{2}{3}$, $Y = \sqrt{\frac{X}{\beta}}$ is *Maxwell*. $Y = \frac{1}{X}$ is *inverted gamma*.				
Logistic(μ, β)	$\frac{1}{\beta} \frac{e^{-\frac{x-\mu}{\beta}}}{\left[1+e^{-\frac{x-\mu}{\beta}}\right]^2};\ \beta > 0$	μ	$\frac{\pi^2\beta^2}{3}$	$e^{\mu t}\Gamma(1+\beta t),\ \lvert t \rvert < \frac{1}{\beta}$
Notes: The cdf is $F(x \mid \mu, \beta) = \frac{1}{1+e^{-\frac{x-\mu}{\beta}}}$.				
Lognormal(μ, σ^2)	$\frac{1}{\sqrt{2\pi}\sigma} \frac{1}{x} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}};\ x > 0, \sigma > 0$	$e^{\mu + \frac{\sigma^2}{2}}$	$e^{2(\mu+\sigma^2)} - e^{2\mu+\sigma^2}$	$EX^n = e^{n\mu + \frac{n^2\sigma^2}{2}}$
Normal(μ, σ^2)	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}};\ \sigma > 0$	μ	σ^2	$e^{\mu t + \frac{\sigma^2 t^2}{2}}$
Pareto(α, β)	$\frac{\beta\alpha^\beta}{x^{\beta+1}};\ x > \alpha,\ \alpha, \beta > 0$	$\frac{\beta\alpha}{\beta-1},\ \beta > 1$	$\frac{\beta\alpha^2}{(\beta-1)^2(\beta-2)},\ \beta > 2$	does not exist
t_ν	$\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{\sqrt{\nu\pi}} \frac{1}{(1+\frac{x^2}{\nu})^{\frac{\nu+1}{2}}}$	$0,\ \nu > 1$	$\frac{\nu}{\nu-2}, \nu > 2$	$EX^n = \frac{\Gamma(\frac{\nu+1}{2})\Gamma(\nu-\frac{n}{2})}{\sqrt{\pi}\Gamma(\frac{\nu}{2})} \nu^{\frac{n}{2}},\ n$ even
Notes: $t_\nu^2 = F_{1,\nu}$.				
Uniform(a, b)	$\frac{1}{b-a},\ a \leq x \leq b$	$\frac{b+a}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{bt}-e^{at}}{(b-a)t}$
Notes: If $a = 0,\ b = 1$, this is special case of beta $(\alpha = \beta = 1)$.				
Weibull(γ, β)	$\frac{\gamma}{\beta} x^{\gamma-1} e^{-\frac{x^\gamma}{\beta}};\ x > 0,\ \gamma, \beta > 0$	$\beta^{\frac{1}{\gamma}}\Gamma(1+\frac{1}{\gamma})$	$\beta^{\frac{2}{\gamma}} \left[\Gamma(1+\frac{2}{\gamma}) - \Gamma^2(1+\frac{1}{\gamma})\right]$	$EX^n = \beta^{\frac{n}{\gamma}}\Gamma(1+\frac{n}{\gamma})$
Notes: The mgf only exists for $\gamma \geq 1$.				

More fun with Exponentials

Find the variance of the exponential.

More fun with Exponentials

Find the variance of the exponential.

It's on the prior slide's tables. Nailed it!

More fun with Exponentials

Find the variance of the exponential.
OR we can compute $E[(X - \mu_x)^2]$

Daily Recap

Today we learned

1. Wrapup on distributions and populations

Moving forward:

- The normal distribution, with its nb day Friday

Next time in lecture:

- Wrap-up and some more examples on populations.