CSCI 3022 Intro to Data Science
Distributions Wrapup and Normals

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

- Exam pushed to Friday of next week
- Practicum posted laster this week!

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: $\displaystyle\sum_{x\in\Omega} xf(x)$; Continuous: $\displaystyle\int_{x\in\Omega} x \cdot f(x)\, dx$
2. Expected value of a function $g(X)$ of X is:
   $\displaystyle\sum_{x\in\Omega} g(x)f(x)$; $\displaystyle\int_{x\in\Omega} g(x) \cdot f(x)\, dx$
3. $Y = g(X)$ is a *change of variables*.
4. Expectation is **linear:** $E[aX + b] = aE[X] + b$

Variance of a Random Variable

Definition: *Variance:*
For a discrete random variable X with pdf $f(x)$ and mean $E[X] = \mu_X$, 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 $f(x)$ and mean $E[X] = \mu_X$, 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: $\sigma = \sqrt{\sigma^2}$

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 $Var[aX + b]$?

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

Calculating Variance

When tasked with computing Variance sums/integrals, it is often a little tedious to compute

$$Var[x] = \quad \sum_x (x - E[x])^2 f(x) \quad \text{or} \quad \sum_x \int (x - E[x])^2 f(x)\, dx$$

.

Calculating Variance

When tasked with computing Variance sums/integrals, it is often a little tedious to compute

$$Var[x] = \sum_x (x - E[x])^2 f(x) \quad \text{or} \quad \sum_x \int (x - E[x])^2 f(x)\, dx$$

.

Important Formula: $Var[X] = E[X^2] - E[X]^2$

Proof:

Calculating Variance

When tasked with computing Variance sums/integrals, it is often a little tedious to compute

$$Var[x] = \sum_x (x - E[x])^2 f(x) \quad \text{or} \quad \sum_x \int (x - E[x])^2 f(x)\, dx$$

.

Important Formula: $Var[X] = E[X^2] - E[X]^2$

Proof:

$$\begin{aligned} Var[X] = E[(X - E[X])^2] &\overbrace{=}^{foil} E[X^2 - 2XE[X] + E[X]^2] \\ &\overbrace{=}^{linear} E[X^2] - E[2XE[X]] + E[E[X]^2] \\ &\overbrace{=}^{E[X]\, non-random} E[X^2] - 2E[X]E[X] + E[X]^2 \\ &\overbrace{=}^{simplify} E[X^2] - E[X]^2 \end{aligned}$$

Calculating Variance

This can help a lot! Note that

$$E[X^2] = \sum x^2 f(x) \quad \text{and} \quad \sum_x \int x^2 f(x)\, dx$$

look like a very similar mechanical computations to

$$E[X] = \sum x f(x) \quad \text{and} \quad \sum_x \int x^2 f(x)\, dx,$$

so we can reuse a lot of work, as we'll always compute $E[x]$ before $Var[X]$ either way!
Important Formula: $Var[X] = E[X^2] - E[X]^2$

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^2_p	$\frac{1}{\Gamma(\frac{p}{2})2^{\frac{p}{2}}}x^{\frac{p}{2}-1}e^{-\frac{x}{2}};\ x>0,\ p\in N$	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)	$\frac{1}{\theta}e^{-\frac{x}{\theta}};\ x\ge 0,\ \theta>0$	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}	$\frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})}\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}}\frac{x^{\frac{\nu_1-2}{2}}}{\left(1+(\frac{\nu_1}{\nu_2})x\right)^{\frac{\nu_1+\nu_2}{2}}};\ x>0$	$\frac{\nu_2}{\nu_2-2},\ \nu_2>2$	$2(\frac{\nu_2}{\nu_2-2})^2\frac{\nu_1+\nu_2-2}{\nu_1(\nu_2-4)},\ \nu_2>4$	$EX^n=\frac{\Gamma(\frac{\nu_1+2n}{2})\Gamma(\frac{\nu_2-2n}{2})}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})}\left(\frac{\nu_2}{\nu_1}\right)^n,\ n<\frac{\nu_2}{2}$
Notes: $F_{\nu_1,\nu_2}=\frac{\chi^2_{\nu_1}/\nu_1}{\chi^2_{\nu_2}/\nu_2}$, where the χ^2s are independent. $F_{1,\nu}=t^2_\nu$.				
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	$e^{\mu t+\frac{\sigma^2t^2}{2}}$
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: $t^2_\nu=F_{1,\nu}$.				
Uniform(a,b)	$\frac{1}{b-a},\ a\le x\le b$	$\frac{b+a}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{bt}-e^{at}}{(b-a)t}$
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 $\gamma\ge 1$.				

More fun with Variances

What are the mean and variance of the continuous uniform distribution? Recall: The pdf is $f(x) = \frac{1}{b-a}$ in $[a, b]$

More fun with Variances

What are the mean and variance of the continuous uniform distribution? Recall: The pdf is $f(x) = \frac{1}{b-a}$ in $[a, b]$ It's on the prior slide's tables. Nailed it!

More fun with Variances

What are the mean and variance of the continuous uniform distribution? Recall: The pdf is $f(x) = \frac{1}{b-a}$ in $[a, b]$

OR we can compute $E[(X - \mu_x)^2]$

More fun with Variances

What are the mean and variance of the continuous uniform distribution? Recall: The pdf is $f(x) = \frac{1}{b-a}$ in $[a, b]$

The mean is $\int_a^b \frac{1}{b-a} x \, dx$, so

$$E[X] = \frac{1}{b-a} \frac{x^2}{2} |_a^b = \frac{1}{b-a} \frac{b^2 - a^2}{2} = \frac{(b-a)(b+a)}{2} = \frac{a+b}{2}$$

More fun with Variances

What are the mean and variance of the continuous uniform distribution? Recall: The pdf is $f(x) = \frac{1}{b-a}$ in $[a, b]$

The mean is $\int_a^b \frac{1}{b-a} x \, dx$, so

$$E[X] = \frac{1}{b-a} \frac{x^2}{2} |_a^b = \frac{1}{b-a} \frac{b^2 - a^2}{2} = \frac{(b-a)(b+a)}{2} = \frac{a+b}{2}$$

The variance is probably easier to compute using the shortcut formula. So lets find

$$E[X^2] = \int_a^b \frac{1}{b-a} x^2 \, dx = \frac{1}{b-a} \frac{x^3}{3} |_a^b$$

$$= \frac{1}{b-a} \frac{b^3 - a^3}{3} = \frac{(b-a)(b^2 + ab + a^2)}{3} = \frac{a^2 + ab + b^2}{3}$$

More fun with Variances

What are the mean and variance of the continuous uniform distribution? Recall: The pdf is $f(x) = \frac{1}{b-a}$ in $[a, b]$

Now we combine these! We have $E[X] = \frac{a+b}{2}$ and $E[X^2] = \frac{a^2+ab+b^2}{3}$, so

$$Var[X] = E[X^2] - E[X]^2 = \frac{a^2 + ab + b^2}{3} - \frac{(a+b)^2}{4}$$

$$= \frac{a^2 - 2ab + b^2}{12} = \frac{(b-a)^2}{12}$$

Another Variance

Find the variance of the face of a fair die.

Another Variance

Find the variance of the face of a fair die.

It's on the prior slide's tables. Nailed it! **OR** we can compute $E[(X - \mu_x)^2]$

The Normal Distribution

The normal distribution (sometimes called the Gaussian distribution) is probably the most important distribution in all of probability and statistics.

Many populations have distributions that can be fit very closely by an appropriate normal (or Gaussian, bell) curve.

Examples: height, weight, and other physical characteristics, scores on various tests, etc.

The Normal Distribution

Definition: *Normal Distribution:*

A continuous r.v. X is said to have a *normal distribution* with parameters _ and __ > 0, if the pdf of X is:

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-1}{2\sigma^2}(x-\mu)^2}$$

Notation: We write ___________

The Normal Distribution

Definition: *Normal Distribution:*

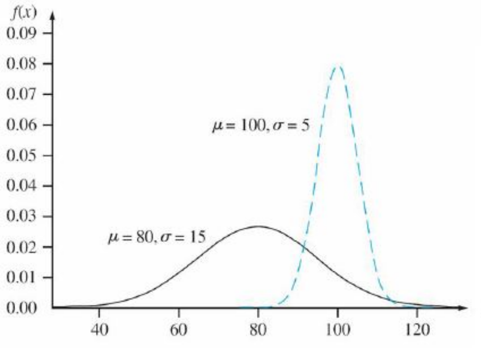
A continuous r.v. X is said to have a *normal distribution* with parameters $\underline{\mu}$ and $\underline{\sigma^2} > 0$, if the pdf of X is:

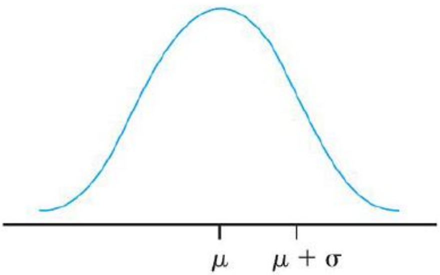
$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-1}{2\sigma^2}(x-\mu)^2}$$

Notation: We write $\underline{X \sim N(\mu\sigma^2)}$

The Normal Distribution

The figure below presents graphs of f for different parameter pairs:





You can play with normals in any statistical software. See for example
https://academo.org/demos/gaussian-distribution/

The Standard Normal Distribution

Definition: *Standard Normal Distribution:*
The normal distribution with parameter values _____ and ______ is called the *standard normal distribution*.

A r.v. with this distribution is called a standard normal random variable and is denoted by Z. Its pdf is:

$$f(z) =$$

The Standard Normal Distribution

Definition: *Standard Normal Distribution:*
The normal distribution with parameter values $\underline{\mu = 0}$ and $\underline{\sigma^2 = 1}$ is called the *standard normal distribution*.

A r.v. with this distribution is called a standard normal random variable and is denoted by Z.
Its pdf is:

$$f(z) =$$

The Standard Normal Distribution

Definition: *Standard Normal Distribution:*
The normal distribution with parameter values $\underline{\mu = 0}$ and $\underline{\sigma^2 = 1}$ is called the *standard normal distribution*.

A r.v. with this distribution is called a standard normal random variable and is denoted by Z.
Its pdf is:

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

The normal cdf

Let's find the cdf of the standard normal distribution!
All we have to to is integrate:

$$\int_{-\infty}^{Z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$$

The normal cdf

Let's find the cdf of the standard normal distribution!
All we have to to is integrate:

$$\int_{-\infty}^{Z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$$

Should we try a substitution? IBP?... this may not go integreat for us.

The normal cdf

Let's find the cdf of the standard normal distribution!
All we have to to is integrate:

$$\int_{-\infty}^{Z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$$

The CDF of the normal distribution has no closed form. But it's really important! So we give it it's own name.

The normal cdf

For a random variable $Z \sim N(0,1)$, the cdf of Z is given by

$$F(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt = \boxed{\Phi(z)}$$

The normal cdf

For a random variable $Z \sim N(0,1)$, the cdf of Z is given by

$$F(z) = P(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt = \boxed{\Phi(z)}$$

Old school statisticians used to carry around giant tables with values of $\Phi(z)$ in them. Actually, many current statisticians do that too, but that's a little silly. We have computers!

The Standard Normal

Note:

1. The standard normal distribution rarely occurs naturally.
2. Instead, it is a reference distribution from which information about other normal distributions can be obtained via a simple formula.
3. These probabilities can then be found "normal tables".
4. This can also be computed with a single command... (`scipy.stats.norm.cdf`, for example)

The Standard Normal

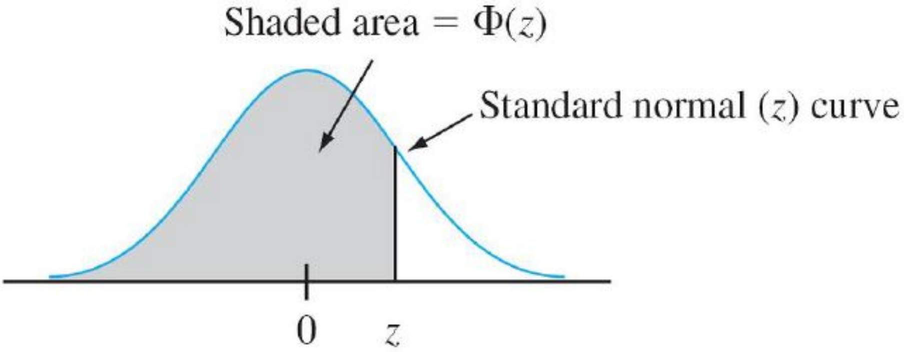
Note:

1. The standard normal distribution rarely occurs naturally.
2. Instead, it is a reference distribution from which information about other normal distributions can be obtained via a simple formula.
3. These probabilities can then be found "normal tables".
4. This can also be computed with a single command... (`scipy.stats.norm.cdf`, for example)

Recall: one example from HW1: if we take a data set, and *subtract the mean* from each of the data values, then we *divide by the standard deviation*, we ended up with a new data set that was mean of 0 and variance/standard deviation of 1. The new data set had the same **shape** as the original, but now it was "centered" at 0 and "scaled" to be of a known (average) spread.

The Standard Distribution

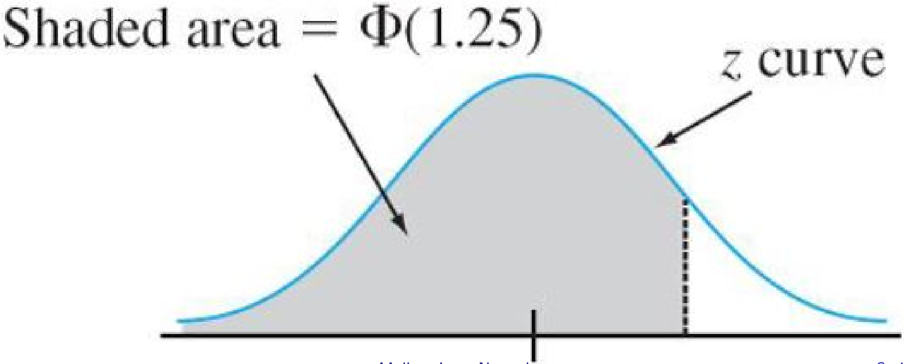
The figure below illustrates the probabilities found in a normal table (such a table can easily be found online):



The Standard Distribution

$P(Z \leq 1.25) = \Phi(1.25)$, a probability that is tabulated in a normal table. What is this probability?

The figure below illustrates this probability:



The Standard Distribution

Some quick examples:

1. $P(Z \geq 1.25)$

2. Why does $P(Z < -1.25) = P(Z > 1.25)$? What is $\Phi(-1.25)$?

3. How do we calculate $P(-.38 \leq Z \leq 1.25)$?

The Standard Distribution

Some quick examples:

1. $P(Z \geq 1.25)$
   It's `1-scipy.stats.norm.cdf(1.25)`. Or as a picture:

2. Why does $P(Z < -1.25) = P(Z > 1.25)$? What is $\Phi(-1.25)$?
   Symmetry! Same as above.

3. How do we calculate $P(-.38 \leq Z \leq 1.25)$?
   As an integral, this is $\displaystyle\int_{-.38}^{1.25} f(z)\, dz$. We could split this into 2:
   $$\int_{-\infty}^{1.25} f(z)\, dz + \int_{-.38}^{-\infty} f(z)\, dz =$$
   $$\Phi(1.25) - \Phi(-.38)$$

Standard Quantiles

The 99th *percentile* of the standard normal distribution is that value of z such that the area under the z curve to the left of the value is 0.99.

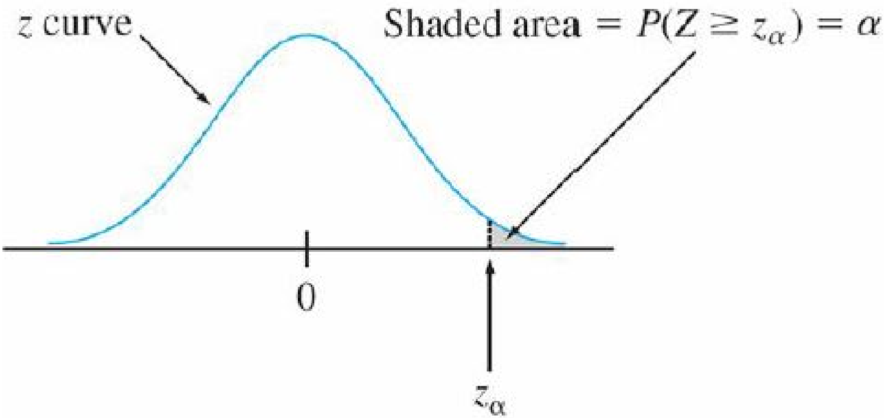
Tables and cdf functions give, for fixed z, the area under the standard normal curve to the left of z; now we have the area and want the value of z.

This is the “inverse” problem to $P(Z \leq z) =?$

How can the table be used for this?

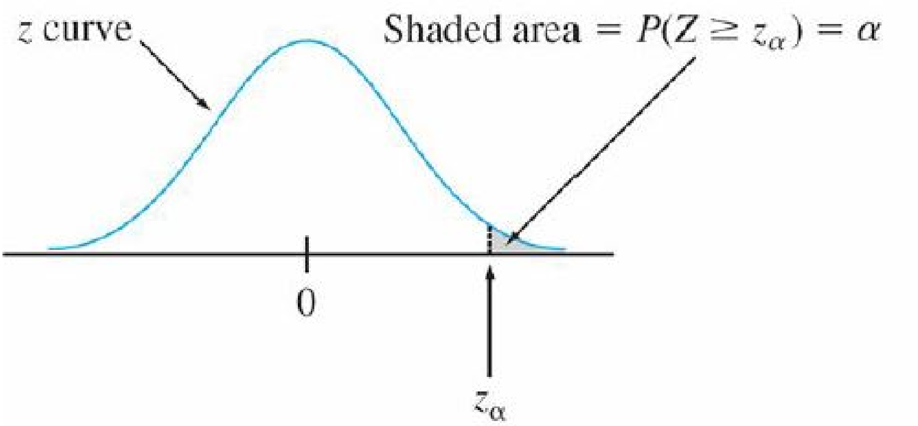
Standard Quantiles

In statistical inference, we need the z values that give certain tail areas under the standard normal curve. There, this notation will be standard: __ will denote the z value for which _ of the area under the z curve lies to the right of __.



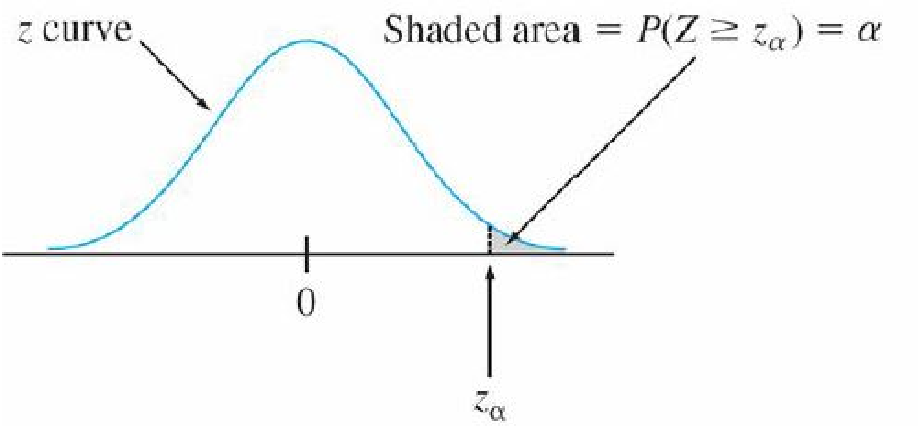
Standard Quantiles

In statistical inference, we need the z values that give certain tail areas under the standard normal curve. There, this notation will be standard: $\underline{z_\alpha}$ will denote the z value for which $\underline{\alpha}$ of the area under the z curve lies to the right of $\underline{z_\alpha}$.



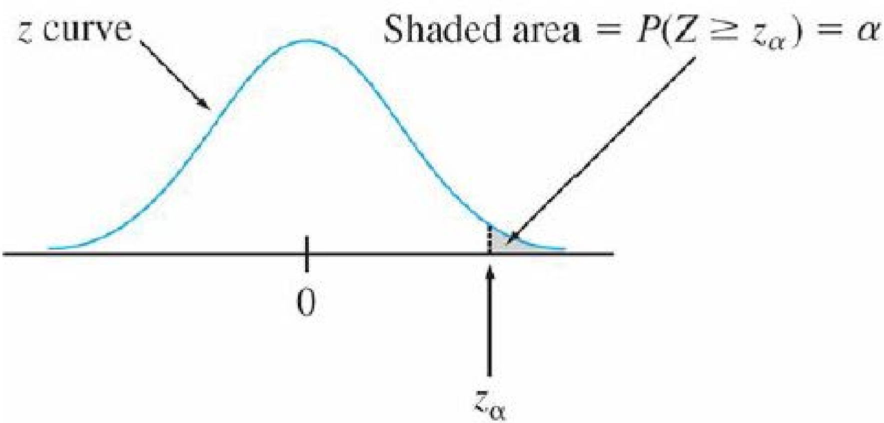
Standard Quantiles

In statistical inference, we need the z values that give certain tail areas under the standard normal curve. There, this notation will be standard: __ will denote the z value for which _ of the area under the z curve lies to the right of __.



Standard Quantiles

In statistical inference, we need the z values that give certain tail areas under the standard normal curve. There, this notation will be standard: $\underline{z_\alpha}$ will denote the z value for which $\underline{\alpha}$ of the area under the z curve lies to the right of $\underline{z_\alpha}$.



Non-Standard Normals

When $\underline{X \sim N(\mu, \sigma^2)}$, probabilities involving X are computed by "standardizing." The standardized variable is:

Proposition: If X has a normal distribution with mean _ and standard deviation _, then

is distributed standard normal.

Non-Standard Normals

When $\underline{X \sim N(\mu, \sigma^2)}$, probabilities involving X are computed by "standardizing." The standardized variable is:

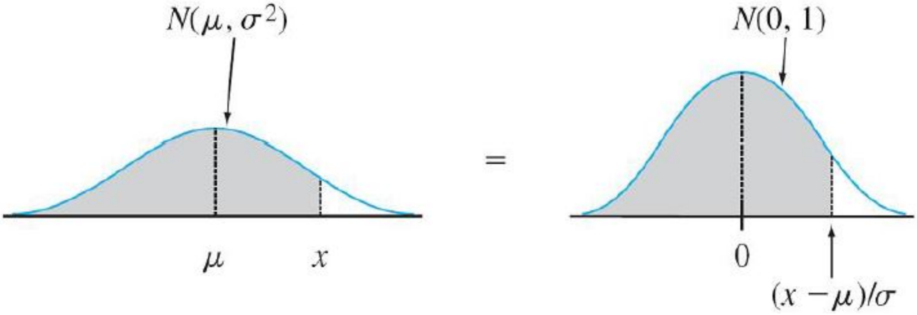
$$Z = \frac{X - \mu}{\sigma}$$

Proposition: If X has a normal distribution with mean $\underline{\mu}$ and standard deviation $\underline{\sigma}$, then

is distributed standard normal.

Non-Standard Normals

Why do we standardize normal random variables?



Equality of nonstandard and standard normal curve areas

Using Normals

Example:
The time that it takes a driver to react to the brake lights on a decelerating vehicle is critical in helping to avoid rear-end collisions.

Research suggests that reaction time for an in-traffic response to a brake signal from standard brake lights can be modeled with a normal distribution having mean value 1.25 sec and standard deviation of 0.46 sec.

What is the probability that reaction time is between 1.00 sec and 1.75 sec?

Solution:

Example: For a normal distribution having mean value 1.25 sec and standard deviation of 0.46 sec.

What is the probability that reaction time is between 1.00 sec and 1.75 sec?

Solution:

Example: For a normal distribution having mean value 1.25 sec and standard deviation of 0.46 sec.

$$X \sim N(1.25, .46)$$

What is the probability that reaction time is between 1.00 sec and 1.75 sec?
We want $P(1 < X < 1.75)$... but we can't compute these probabilities unless the r.v. in the middle of the inequality is *standard* normal. So we normalize!

Solution:

Example: For a normal distribution having mean value 1.25 sec and standard deviation of 0.46 sec.

What is the probability that reaction time is between 1.00 sec and 1.75 sec?
We want $P(1 < X < 1.75)$... but we can't compute these probabilities unless the r.v. in the middle of the inequality is *standard* normal. So we normalize!

$$P(1 < X < 1.75) = P(1 - 1.25 < X - 1.25 < 1.75 - 1.25)$$

$$= P(\frac{-.25}{.46} < \frac{X - 1.25}{.46} < \frac{.5}{.46}) = P(\frac{-.25}{.46} < Z < \frac{.5}{.46})$$

$$= \Phi(\frac{-.25}{.46}) - \Phi(\frac{.5}{.46})$$

Daily Recap

Today we learned

1. Variance and introduced the Normal Distribution

Moving forward:

- nb day Friday!

Next time in lecture:

- Why the Normal matters