Negative Binomial Distribution

You're watching a potentially endless sequence of Bernoulli trials with constant success probability p. Let X = # failures before r^{th} success, r integer ≥ 1.

You can show that X follows the Negative Binomial dist:

its PF is $f(x \mid r, p) = \binom{r+x-1}{x} p^r (1-p)^x$ $I_{\{0,1,2,\ldots\}}(x)$.

with parameters (r, p) $(0 < p < 1)$

The name comes from the fact that, when you watch a sequence of Bernoulli trials with constant (unknown) success probability p unfold, there are two different ways to

estimate p: decide ahead of time to sample n (known constant) success/failure trials, and record the (random) # S of successes you see (from which a reasonable estimate would be $\hat{p}_B = \frac{S}{n}$ ← Binomial).

---

or decide ahead of time that you're going to sample until you've seen s (known constant) successes & record the (random) # of trials N needed to accumulate that many successes (from which a reasonable estimate would be $\hat{p}_{NB} = \frac{s}{N}$ ← Negative Binomial).

Special case of Negative Binomial

Set $r=1$ and record the number X of failures until the first success: X is said to follow the

Geometric (p) distribution, with

pf $f_X(x \mid p) = p(1-p)^x \, I_{\{0,1,\dots\}}(x)$

(parameter p)

Consequence $X_1, \dots, X_r$ IID Geometric(p)

$\rightarrow \sum_{i=1}^{r} X_i \sim$ Negative Binomial (r, p)

This is a direct analogue to the Bernoulli/Binomial story: $X_1, \dots, X_n$ IID Bernoulli(p) $\rightarrow \sum_{i=1}^{n} X_i \sim$ Binomial(n, p).

$X \sim$ Negative Binomial (r, p)

$$\psi_X(t) = \left[\frac{p}{1-(1-p)e^t}\right]^r \quad \text{for } t < \log\left(\frac{1}{1-p}\right)$$

from which $E(X) = \dfrac{r(1-p)}{p}$, $V(X) = \dfrac{r(1-p)}{p^2}$

Consequence $X \sim$ Geometric $(p) \rightarrow$

$\left\{\begin{matrix} k \\ t \end{matrix}\right\}$ both non-negative integers

$$P(X = k+t \mid X \geq k) = P(X = t)$$

this is called the memoryless property of the Geometric distribution, and it turns out that this is the only

discrete distribution with this property.

F F F F F S

X

Y

X = # failures until first success = 5 (here)

Y = # failures, starting at trial $(k+1 = 2)$ here, until next success (= 4 here)

Then Y has the same dist. as X and is independent of what happened on the first k trials, i.e., "the process has no memory".

Case 2: Important Continuous Distributions

Normal (Gaussian) Distribution

$X \sim \text{Normal}(\mu, \sigma^2)$ mean μ, variance $0 < \sigma^2 < \infty$

PDF =

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

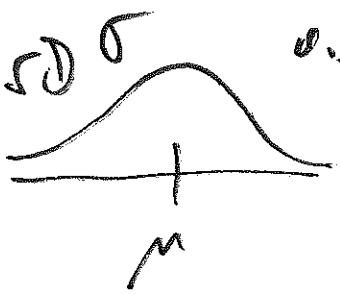
The Normal dist. is the single most important dist. in all of probability & statistics, mainly for 2 reasons: (1) many observable random processes have dist. shapes that are close to the "bell curve" (Normal PDF), and (2) the Central Limit Theorem (CLT), which we'll examine soon.

Properties of the Normal Dist.

$X \sim \text{Normal}(\mu, \sigma^2)$ | $E(X) = \mu$

$V(X) = \sigma^2$, $SD(X) = \sigma$

$$\psi_X(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$$



dist. of X

$\left(\begin{array}{c}\text{center of symmetry} \\ \text{mean} \\ \text{median} \\ \text{mode}\end{array}\right)$ all $= \mu$

Consequences ① $X \sim \text{Normal}(\mu, \sigma^2)$,

$Y = aX + b$, $\binom{a \neq 0}{b}$ fixed constants $\rightarrow$

$Y \sim \text{Normal}(a\mu + b, a^2\sigma^2)$.

In other words, Normality is preserved under linear transformation

---

Def.

The Normal dist. with mean $\mu = 0$ and SD $\sigma = 1$ is called the <u>standard</u> <u>Normal dist</u>.

The PDF of $X \sim \text{Normal}(0, 1)$ is

$$\phi_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$ and its

phi (lower-case)

CDF is $\Phi(x) = \int_{-\infty}^{x} \phi_X(t)\, dt$

upper case Phi

It turns out that e^{-cx^2} $(c>0)$ has no anti-derivative in closed form, so $\Phi(x)$ cannot be summarized in a formula; instead it's approximated by numerical integration (see p 861 in DS).

Consequences, continued

② Because the Normal PDF (for all $x \in \mathbb{R}$) is symmetric, $\Phi(-x) = 1 - \Phi(x)$

and $\Phi^{-1}(p) = -\Phi^{-1}(1-p)$ (for all $0<p<1$)

---

③ $X \sim \text{Normal}(\mu, \sigma^2) \rightarrow Z = \frac{X-\mu}{\sigma} \sim N(0,1)$,

so that $F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$

and $F_X^{-1}(p) = \mu + \sigma\,\Phi^{-1}(p)$

Empirical Rule

Part 1 Start at the mean μ of a distribution and go 1 SD σ either way: you will find (about $\frac{2}{3}$) 68% of the probability in the interval $(\mu \pm 1\sigma)$

Part 2 Ditto 2 SDs either way: $(\mu \pm 2\sigma)$ capturing (about most 95%) of the probability

Part 3 Ditto 3 SDs either way: $(\mu \pm 3\sigma)$ capturing almost all 99.7% of the probability

This Rule is <u>exact</u> for all Normal dists & is a surprisingly

good approximation for many other distributions. This permits an easy trick that's helpful in computing Normal probabilities.

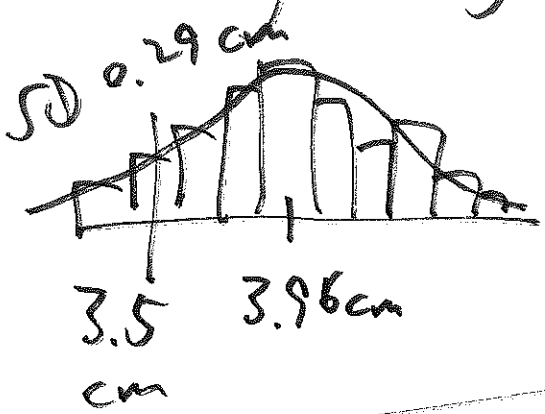
Example: You have a random sample of $n = 103$ immature monarch butterflies, and you measure their wing lengths:

y = wing length (cm)

$$\begin{bmatrix} y_1 = 4.1 \\ y_2 = 3.3 \\ \vdots \\ y_n = 4.7 \end{bmatrix} \quad n = 103$$

mean $\bar{y} = 3.96$ cm

SD $s = 0.29$ cm



histogram of wing length

Q: About what % of the sampled butterflies had wing length ≤ 3.5 cm?

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

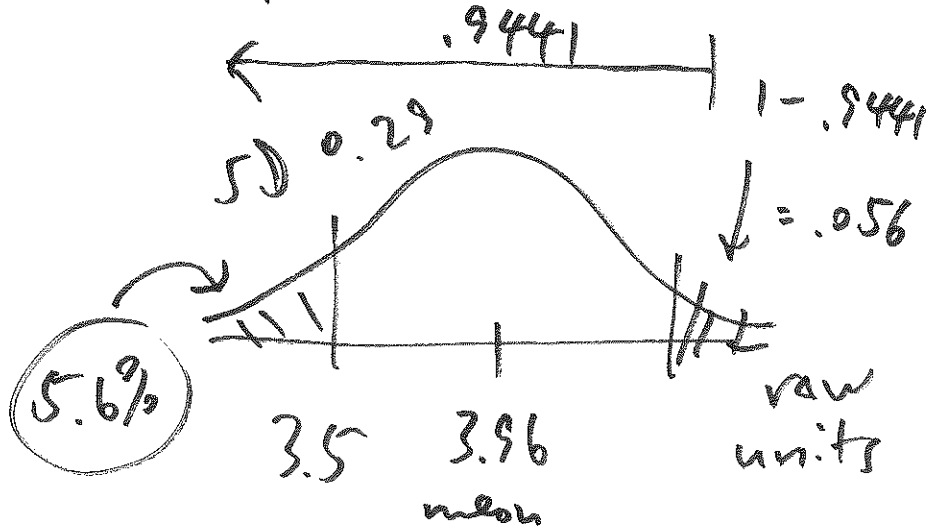
sample mean sample SD

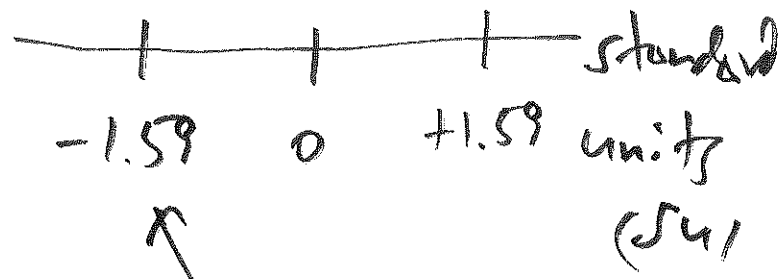
sorted y

3.2	4	
3.3		
⋮		
3.5	8	
3.5		
3.5		103
3.5		
3.6		
⋮		
4.7		

$\boxed{A_1}$ (exact) $\frac{8}{103} \doteq 7.8\%$ (263)

$\boxed{A_2}$ (approximate)





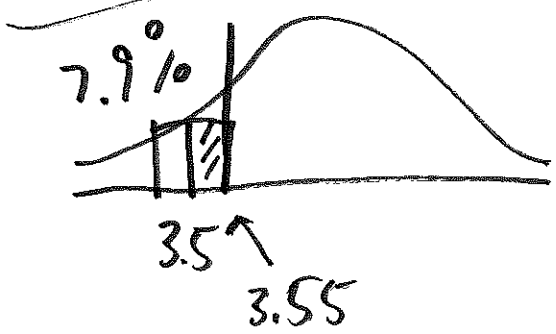
$$\frac{3.5 - 3.96}{0.29}$$

converting to SU for data:

$$z = \frac{y - \bar{y}}{s} = SU$$

for random variables

$$Z_Y = \frac{Y - \mu}{\sigma} = SU$$



keeping track of histogram bar edges: continuity correction

More consequences

(4) $X_1, \ldots, X_k$ independent, $X_i \sim \text{Normal}(\mu_i, \sigma_i^2)$

$$\rightarrow \sum_{i=1}^{k} X_i \sim \text{Normal}\left(\sum_{i=1}^{k} \mu_i, \sum_{i=1}^{k} \sigma_i^2\right)$$

nice additive property

this is why Normal dists are indexed by variance rather than SD.

Notation: $\text{Normal}(\mu, \sigma^2) \triangleq N(\mu, \sigma^2)$

Example: Population of adult U.S. women: height follows $N(\mu = 65.0\text{in}, \sigma^2 = 3.2^2\text{in}^2)$ dist.

$(\sigma = 3.2\text{ in})$

Pop. of adult U.S. men: height follows $N(\mu = 69.5\text{ in}, \sigma^2 = 3.3^2\text{ in}^2)$ dist.

1 woman chosen at random, height $\underline{W}$; (265)

1 man chosen at random (independently), height $\underline{M}$; P(woman taller than man)

$= P(\underline{W} > \underline{M})$

$= ?$

Define $D = W - \underline{M}$

By consequence (4), $D \sim N(65 - 69.5 = -4.5 \text{ in},$

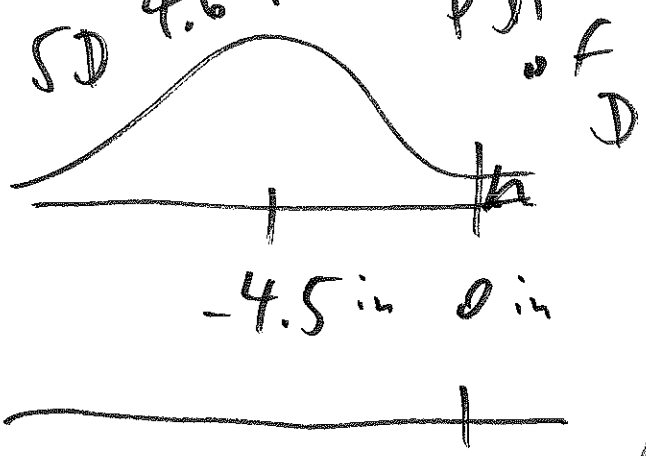
$3.2^2 + 3.3^2 = 21.1 \text{ in}^2)$

$P(\underline{W} > \underline{M}) = P(D > 0)$

SD $\sqrt{21.1 \text{ in}^2}$

$= 4.6$ in

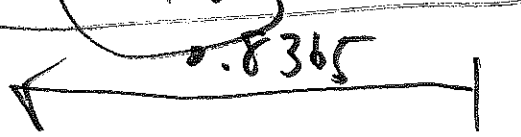
SD 4.6 in



Convert to SU:

$$\frac{0 - (-4.5)}{4.6} = +0.98$$

DS table



So $P(\underline{W} > \underline{M}) \doteq 16\%$

Def rv $X_1, \ldots, X_n \to$ <u>sample mean</u> of $(X_1, \ldots, X_n)$ is $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

Consequences, continued

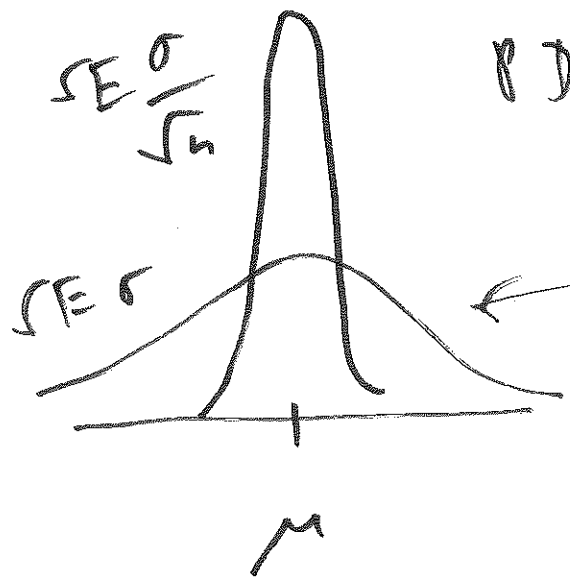
(5) $\left\{ X_i \overset{IID}{\sim} N(\mu, \sigma^2) \quad (i = 1, \ldots, n) \right\}$

$\to \quad \bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ so $SD(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}$

Because $E(\bar{X}_n) = \mu$, $\bar{X}_n$ is an <u>unbiased</u> <u>estimator</u> of μ

Def. In frequentist statistics, the standard deviation (SD) of an estimator $\hat{\theta}$ (rv) of a parameter θ is called the <u>standard error</u> $SE(\hat{\theta})$ of $\hat{\theta}$.

So if you use $\overline{X}_n$ as an estimate of μ, $SE(\overline{X}_n) = \frac{\sigma}{\sqrt{n}} \to 0$ as $n \to \infty$



This is the basis of most of frequentist statistical inference

As $n\uparrow$, $\overline{X}_n$ gets better as an estimate of μ, at a $\sqrt{n}$ rate; this is called the <u>square root law</u>.

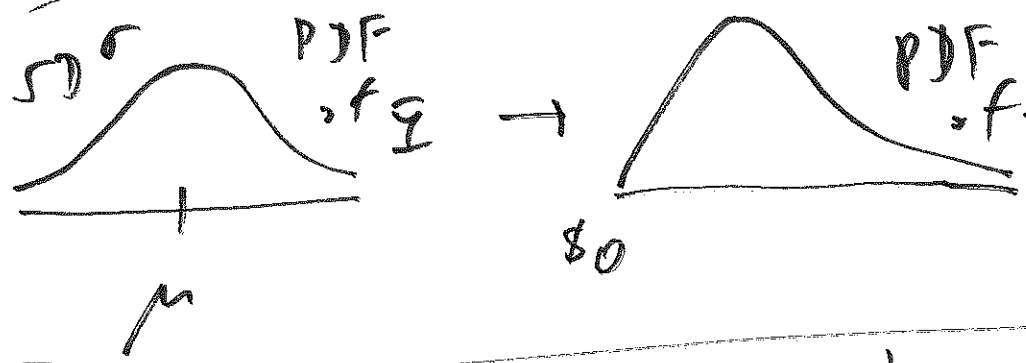
Unfortunately, this means that to cut the $SE(\overline{X}_n)$ in half, you have to <u>quadruple</u> the sample size.

Log Normal Distribution (This distribution is mis-named; it should be called the <u>Exponential Normal</u> distribution, but we're stuck with a bad name.)

Def.

$X > 0$

If $Y = \log(X) \sim N(\mu, \sigma^2)$, people say that $X \sim$ <u>Log Normal</u> (μ, σ^2).



$X \sim \text{LogNormal}(\mu, \sigma^2)$

- $Y = \log(X) \sim N(\mu, \sigma^2)$

Can get MGF of X from MGF of Y

MGF of X is $\psi_X(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$

But by definition

$$\psi_X(t) = E\left(e^{tX}\right) = E\left(e^{t \log X}\right)$$

$$= E\left(X^t\right),$$

so we can read the moments of X directly from the MGF of X

$$E(X) = \psi_X(1)$$

$$= \exp\left(\mu + \frac{\sigma^2}{2}\right)$$

$$V(X) = \psi_X(2) - \left[\psi_X(1)\right]^2$$

$$= \exp\left(2\mu + \sigma^2\right)\left[e^{\sigma^2} - 1\right].$$

Famous example

Pricing stock options, continued

I shape of a stock, current price S_0 (known constant). Heroic assumption: price

u time units in the future will be

$$S_u = S_0 e^{Z_u}, \quad Z_u \sim N(\mu u, \sigma^2 u).$$

Can write $S_0 e^{Z_u} = e^{Z_u + \log(S_0)}$. Now

$$\left[Z_u + \log(S_0)\right] \sim N(\mu u + \log(S_0), \sigma^2 u),$$

so $S_u \sim \text{LogNormal}(\mu u + \log(S_0), \sigma^2 u)$.

Consider a single time horizon u;

heroic assumption reworded

$$S_u = S_0 \exp\left[\mu u + (\sigma\sqrt{u}) \cdot Z_1\right], \quad Z_1 \sim N(0,1)$$

we need to price the option to buy 1 share of this stock for price q at time u.

Use risk-neutral pricing as in the previous discussion: force present value $E(S_u) \triangleq S_0$.

Let time scale of u be in years; let risk-free (continuous-compounding) interest rate be r/year;

then present value of $E(S_u)$ is $e^{-ru} \cdot E(S_u)$.

But by heroic lognormal assumption,

$$E(S_u) = S_0 \exp\left(\mu u + \frac{\sigma^2 u}{2}\right)$$

So set S_0 equal to

$$e^{-ru} S_0 \exp\left(\mu u + \frac{\sigma^2 u}{2}\right)$$

result is $\mu = r - \frac{\sigma^2}{2}$

for risk-neutral pricing.

Value of option at time u will be $h(S_u)$, where $h(s) = \begin{cases} s-q & \text{if } s > q \\ 0 & \text{else} \end{cases}$.

With $\mu = r - \frac{\sigma^2}{2}$, $h(S_u) > 0$ iff

$$Z_1 > \frac{\log\left(\frac{q}{S_0}\right) - \left(r - \frac{\sigma^2}{2}\right)u}{\sigma\sqrt{u}} \triangleq c$$

Now a nasty integral arises: risk-neutral price of option is the present value of $E[h(S_u)]$, which is

$$e^{-ru} E[h(S_u)] = e^{-ru} \int_c^\infty \left[S_0 e^{(r-\frac{\sigma^2}{2})u + \sigma z\sqrt{u}} - q\right] \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz.$$

Careful calculation reveals the (famous) formula

$S_0 \Phi(\sigma\sqrt{u} - c) - q e^{-ru} \Phi(-c)$ is

the risk-neutral price of the option,

where $c = \dfrac{\log\left(\frac{q}{S_0}\right) - \left(r - \frac{\sigma^2}{2}\right)u}{\sigma\sqrt{u}}$.

(Black-Scholes formula)

This formula was derived in 1973 by

(American economist) → Fischer Black (1938-1995) and Myron Scholes (1941-) (age 57 throat cancer)

↑ won Nobel prize in Economics for this work in 1997, together with Robert Merton (1944-)

(American economist)

Gamma Distribution

$(\alpha, \beta > 0)$ X has the Gamma dist. with parameters (α, β), with $X \sim \Gamma(\alpha, \beta)$ or $X \sim \text{Gamma}(\alpha, \beta)$ →

(Canadian-American economist)

X continuous on $(0, \infty)$ with

PDF $f_X(x \mid \alpha, \beta) = \dfrac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} I(x>0)$

(*)

α is called a <u>shape parameter</u> in the $\Gamma(\alpha, \beta)$ family because it governs things like skewness of the dist.

β is related to the <u>scale</u> of the distribution, which measures how spread out the dist. is

$\Gamma(\alpha)$ is the Gamma function, invented to deal with integrals of functions like (*) above:

$$\Gamma(\alpha) \triangleq \int_0^\infty x^{\alpha-1} e^{-x} \, dx$$

has no anti-derivative in closed form

$\Gamma(\alpha)$ turns out to be a continuous generalization of the factorial function, because $\begin{pmatrix} n \text{ positive} \\ \text{integer} \end{pmatrix} \to \Gamma(n) = (n-1)!$

$\Gamma(\alpha) \to \infty$ really quickly as $\alpha \to \infty$, so it's better to evaluate the Gamma PDF on the log scale and then exponentiate:

$$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} = \exp\left[\begin{array}{l} \alpha \ln(\beta) - \ln\Gamma(\alpha) \\ + (\alpha-1)\ln(x) - \beta x \end{array}\right]$$

Another way to tame $\Gamma(x)$ is with a Stirling's approximation: $\Gamma(x) \doteq \sqrt{2\pi}\, x^{x-\frac{1}{2}} e^{-x}$ for large x

so that $\ln \Gamma(x) \doteq \frac{1}{2}\ln(2\pi) + (x-\frac{1}{2})\ln x - x$.

$X \sim \Gamma(\alpha, \beta)$

$\psi_X(t) = \left(1 - \frac{t}{\beta}\right)^{-\alpha}$ for $t < \beta$

so $E(X) = \frac{\alpha}{\beta}$

and $V(X) = \frac{\alpha}{\beta^2}$

$SD(X) = \frac{\sqrt{\alpha}}{\beta}$

Alternative expression

$\psi_X(t) = \left(\frac{\beta}{\beta - t}\right)^{\alpha}$ for $t < \beta$

Special case of $\Gamma(\alpha, \beta)$

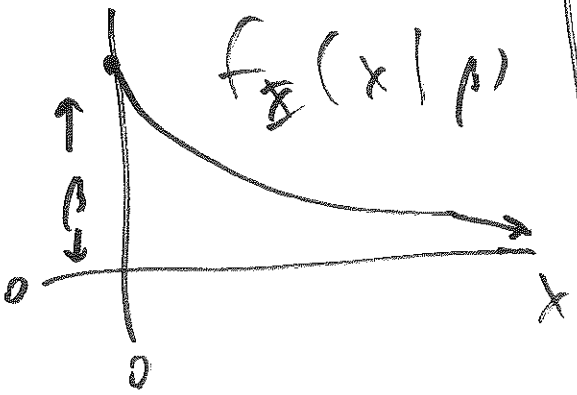
With $\alpha = 1$ the PDF is

$f_X(x \mid \beta) = \beta e^{-\beta x} I(x > 0)$

But this is just our old friend the <u>Exponential</u> <u>distribution</u>.

$X \sim \text{Exponential}(\beta)$ | $\psi_X(t) = \dfrac{\beta}{\beta - t}, \quad t < \beta$

$E(X) = \dfrac{1}{\beta}$ | $V(X) = \dfrac{1}{\beta^2}$ | $SD(X) = \dfrac{1}{\beta}$



~~Notice that the Exponential distribution has TMR equal to 1; this suggests it's related somehow to the Poisson dist.~~

Theorem Suppose that arrivals (events) occur according to a Poisson process with rate β per unit time. Set Z_k = time until k^{th} arrival, $k = 1, 2, \ldots$ and define

$$Y_1 = Z_1$$
$$Y_2 = Z_2 - Z_1$$
$$\ldots \quad Y_k = Z_k - Z_{k-1} \text{ for } k = 2, 3, \ldots$$

the Y_i are called the <u>inter-arrival</u> times.

Then it turns out that $Y_i \overset{IID}{\sim}$ Exponential(β)

The Exponential dist. is also related to the Geometric dist., in that they both have a <u>memoryless property</u>

Theorem

$X \sim$ Exponential(β) ; $t > 0$, $h > 0$

$$\rightarrow P(X \geq t+h \mid X \geq t) = P(X \geq h)$$

Example X = time from initial use until a manufactured product fails

$F_X(x) = P(X \leq x)$

$1 - F_X(x) = P(X > x)$

$= P($"system surviving" at least to time $x)$

For this reason, $1 - F_X(x)$ is called the <u>survival function</u> $S_X(x) = 1 - F_X(x)$ in medicine and the <u>reliability function</u> $R_X(x) = 1 - F_X(x)$ in engineering.

Earlier we showed that for $X \sim \text{Exponential}(\beta)$ $\Rightarrow$ $F_X(x) = 1 - e^{-\beta x}$ for $x > 0$

So $S_X(x) = R_X(x) = e^{-\beta x}$ for this dist.

The <u>instantaneous failure rate</u> or <u>hazard rate</u> function is defined to be $H_X(x) = \dfrac{f_X(x)}{S_X(x)} = \dfrac{f_X(x)}{R_X(x)}$

This gives $P(\text{failure in interval } (x, x+\epsilon) \mid \text{survival to time } x)$ for small ϵ

Notice that if $X \sim \text{Exponential}(\beta)$

then $H_X(x) = \dfrac{\beta e^{-\beta x}}{e^{-\beta x}} = \beta \left(\dfrac{\text{Constant in}}{x} \right)$

The Exponential is the only failure rate distribution with <u>constant hazard</u>. Returning to the earlier result that $X \sim \text{Exponential}(\beta)$,

$\rightarrow P(X \geq t+h \mid X \geq t) = P(X \geq h)$,

---

for all $t > 0$, $h > 0$ this says that if the product has survived to time t, the chance it will survive to time $(t+h)$ is the same as the original chance of surviving from time 0 to time h; i.e., "the system doesn't remember how long it's survived"

Consequences ① $X_i \overset{IID}{\sim} \text{Exponential}(\beta)$ $(i=1,\dots,n)$,

then

$$Y_1 = \min(X_1,\dots,X_n) \sim \text{Exponential}(n\beta).$$

Beta Distribution $\alpha,\beta>0$ $X \sim \text{Beta}(\alpha,\beta) \leftrightarrow$

$$f_X(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \cdot I(0<x<1)$$

The name comes from the normalizing constant: the function $x^{\alpha-1}(1-x)^{\beta-1}$ has no closed-form anti-derivative, so people just made a

---

Definition For all $\alpha>0$, $\beta>0$ $B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx$

beta function

Can show that $B(\alpha,\beta) = \dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.

(α,β) jointly control the shape of the Beta(α,β) dist.

$X \sim \text{Beta}(\alpha,\beta)$

$$\psi_X(t) = 1 + \sum_{k=1}^{\infty}\left(\prod_{r=0}^{k-1}\frac{\alpha+r}{\alpha+\beta+r}\right)\frac{t^k}{k!}$$

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

$$V(X) = \left(\frac{\alpha}{\alpha+\beta}\right)\left(\frac{\beta}{\alpha+\beta}\right)\left(\frac{1}{\alpha+\beta+1}\right)$$

(19 Aug 16)

Example (Castaneda v. Partida continued)