and $P(X = x \mid Y = y) = \binom{y}{x} p^x (1-p)^{y-x}$

Notice that if $X = x$, $Y \geq x$ because the actual number of oocysts (Y) has to be at least as large as the number of oocysts detected (X). After a careful calculation you get:

$$f_X(x) = \sum_{y=x}^{\infty} \binom{y}{x} p^x (1-p)^{y-x} \frac{(\lambda t)^y e^{-\lambda t}}{y!}$$

$$= \frac{e^{-p\lambda t} (p\lambda t)^x}{x!}$$

i.e., $X \sim \text{Poisson}(p\lambda t)$: losing a proportion $(1-p)$ of the oocysts to faulty counting just lowers the rate of the Poisson process from λ/liter to $\lambda \cdot p$/liter (makes excellent sense).

In practice oocysts are hard to detect: p is small (not far from 0). $\boxed{Q:}$ How much water (t liters) do you need to filter to achieve $P(\text{at least 1 oocyte detected}) \geq 1-\alpha$ for small α? $\boxed{A:}$ Not hard to work out

$$P(\text{at least 1 detected}) = 1 - P(\text{none detected})$$

$$= 1 - P(X = 0) = 1 - e^{-p\lambda t} \geq 1-\alpha$$

$$\leftrightarrow \alpha \geq e^{-p\lambda t} \leftrightarrow \ln\alpha \geq -p\lambda t \leftrightarrow$$

$$\boxed{t \geq -\frac{\ln\alpha}{p\lambda}}$$

$\boxed{\text{Example}}$ $\alpha = .01$, $p = 0.1$, $\lambda = 0.2 / \text{liter}$ (1 per 5 liters)

minimum sickness level

to achieve prob 99%, t has to be at least $\underline{230.3}$ liters.

253

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 what's called the <u>Negative Binomial</u> dist.: its PMF is $f_X(x \mid r, p) = \binom{r+x-1}{x} p^r (1-p)^x \, I_{\{0,1,2,\ldots\}}(x)$.

<u>with parameters</u> (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 255

Geometric (p) distribution, with

support of X

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

(parameter p)

Consequence | $X_1, \dots, X_n$ IID Geometric(p)

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

$\underline{X} \sim$ Negative Binomial (r, p)

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

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

Consequence: $\underline{X} \sim$ Geometric $(p) \rightarrow$

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

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

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

$$\underbrace{F\ \underbrace{F\ F\ F\ F\ S}_{Y}}_{X}$$

X = # failures until <u>first</u> success = 5 (here)

Y = # failures, starting at trial $\begin{pmatrix} k+1 \\ =2 \end{pmatrix}$ here, until <u>next</u> success (= 4 here)

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

Case 2: Important Continuous Distributions

Normal (Gaussian) Distribution

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

PDF:

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

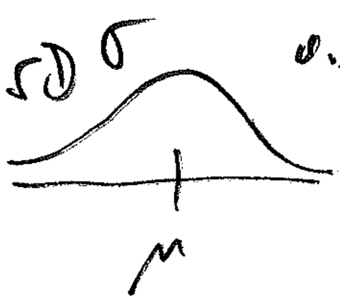
The Normal dist. is the single most important dist. in all of probability & statistics, mainly for 2 reasons: (1) many observable random processes have dist. shapes that are close to the "bell curve" (Normal PDF), and (2) the <u>Central</u> <u>Limit</u> <u>Theorem</u> (CLT), which we'll examine soon.

Properties of the Normal Dist.

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

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

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



dist. of X

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

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

$Y = aX + b$, $\left(\begin{matrix} a \neq 0 \\ b \end{matrix}\right)$ fixed constants $\longrightarrow$

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

In other words, Normality is preserved under linear transformation

Def. The Normal dist. with mean $\mu = 0$ and SD $\sigma = 1$ is called the standard Normal dist.

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

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

phi (lower-case)

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

uppercase Phi Φ

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).

260

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: $(\mu \pm 2\sigma)$ captures (about 95%) of the probability

most

Part 3 Ditto 3 SDs either way: $(\mu \pm 3\sigma)$ captures 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.

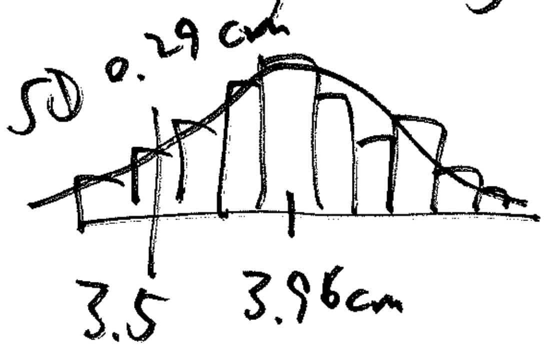
Example: You have a random sample of $n = 103$ immature monarch butterflies, and you measure their wing lengths:

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

263

sorted y

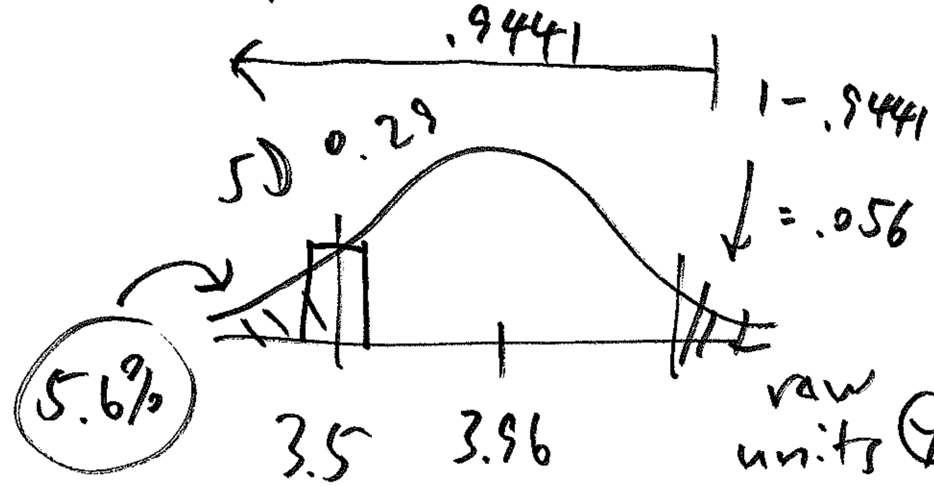
3.2
3.3
⋮
3.5
3.5
3.5
3.5

3.6
⋮
4.7

4 8 103

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

$\boxed{A_2}$ (approximate)



standard units (z) (SU)

−1.59 0 +1.59

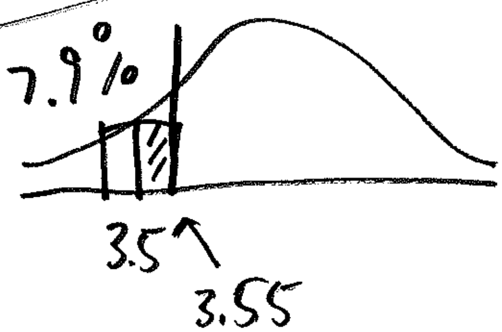
$$\frac{3.5 - 3.96}{0.29}$$

~~converting~~ to SU for data:

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

for random variables

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



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

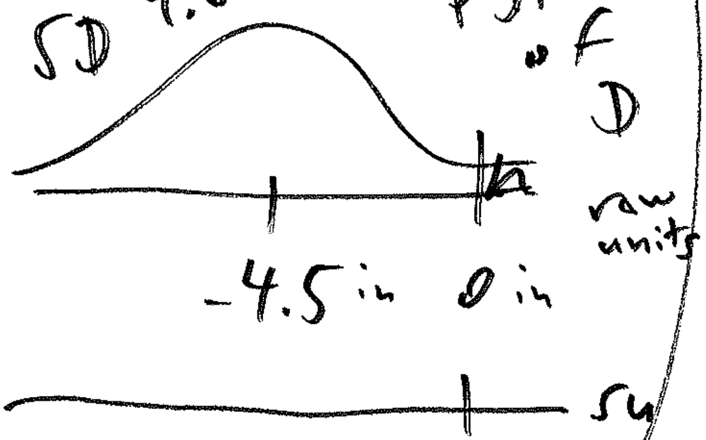
By consequence (4), $D \sim N\left(65 - 69.5 = -4.5 \text{ in}, \; 3.2^2 + 3.3^2 = 21.1 \text{ in}^2\right)$

SD $\sqrt{21.1 \text{ in}^2} = 4.6$ in

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

SD 4.6 in

PDF of D



Convert to su:

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

DS table

.8365

.1635

so $P(\underline{W} > \underline{M}) \doteq 16\%$ (about 1 in 6)

(28 May 19)

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

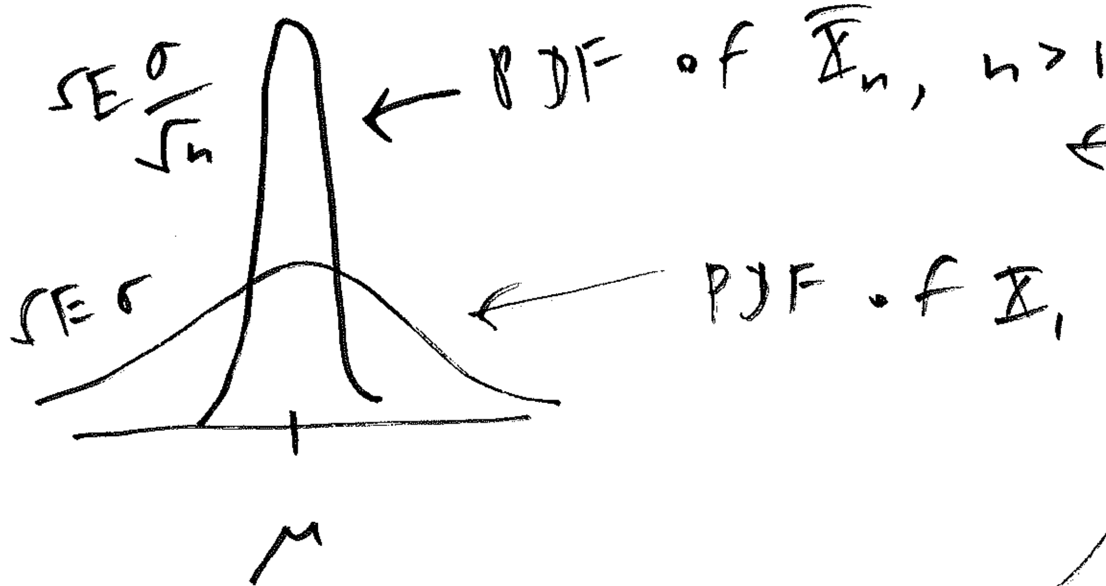
(5) $\left\{ X_i \overset{IID}{\sim} N(\mu, \sigma^2) \quad (i = 1, \ldots, n) \right\}$

$\to \quad \bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ 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> $\widehat{SE}(\hat{\theta})$ of $\hat{\theta}$.

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



this is the basis of most of frequentist statistical inference

---

As $n \uparrow$, $\bar{X}_n$ gets better as an estimate of μ, at a $\frac{1}{\sqrt{n}}$ rate;

this is called the Square Root Law.

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

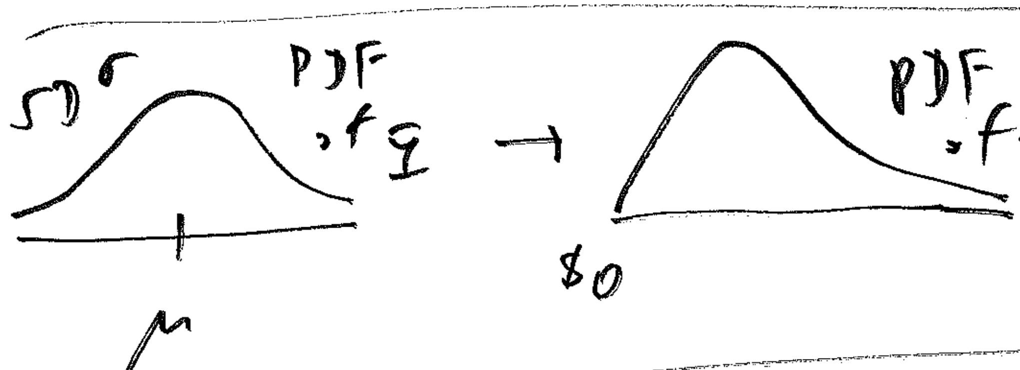
Log Normal Distribution

(This distribution is mis-named (1900); it should be called the Exponential-Normal distribution, but we're stuck with a bad name.)

Def.

$X > 0$

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



$\rightarrow X = e^{Y} = e^{N(\mu, \sigma^2)}$

$X \sim$ LogNormal (μ, σ^2)

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

Can get MGF of X from MGF of Y

MGF of I is $\psi_I(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$

But by definition

$\psi_X(t) = E(e^{tI}) = E(e^{t \log X})$

$= E(X^t)$, so we can

read the moments of X directly from the MGF of I

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

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

$V(X) = \psi_I(2) - [\psi_I(1)]^2$

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

Famous Case Study ~~example~~

Pricing stock options, continued

1 share of a stock, current price S_0 (known constant). <u>Heroic assumption</u>: price

u time units in the future will be

$$S'_u = S'_0 \, e^{Z_{4u}}, \quad Z_{4u} \sim N(\mu u, \sigma^2 u).$$

Can write $S'_0 \, e^{Z_{4u}} = e^{Z_{4u} + \log(S'_0)}$. Now

$$\left[Z_{4u} + \log(S'_0) \right] \sim N\left(\mu u + \log(S'_0), \sigma^2 u \right),$$

so $S'_u \sim \text{LogNormal}\left[\mu u + \log(S'_0), \sigma^2 u \right]$.

Consider a single time horizon u;

heroic assumption reworded

$$S'_u = S_0 \exp\left[\mu u + (\sigma \sqrt{u}) \cdot Z_1 \right],$$

$$Z_1 \sim N(0, 1)$$

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 $E(S_u) \stackrel{\Delta}{=} S_0$. Let time scale of u be in years; let the risk-free (continuous-compounding) interest rate be r / year;

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

But by heroic lognormal assumption,

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

S_0 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

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

Careful calculation reveals the following (famous) formula:

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

the risk-neutral price of the option,

where $c = \dfrac{\log\left(\frac{q}{S_0}\right)-\left(r-\frac{\sigma^2}{2}\right)u}{\sigma\sqrt{u}}$ ← This formula was derived in 1973 by

(Black-Scholes formula) "shoals"

(American economist) → Fischer Black (1938-1995) died age 57 (throat cancer) and

(Canadian-American economist) → Myron Scholes (1941-) ↑ won Nobel prize in Economics for this work in 1997, together with Robert

(American economist) → Merton (1944- 2003)

Gamma Distribution

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

PDF $f_X(x \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \underbrace{I(x>0)}_{\text{support of } X}$ (*)

α is called a _shape parameter_ in the $\Gamma(\alpha, \beta)$ family because it governs things like skewness of the dist. β is related to the _scale_ 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 $\begin{pmatrix} n \text{ positive} \\ \text{integer} \end{pmatrix} \rightarrow \Gamma(n) = (n-1)!$

$\Gamma(\alpha) \rightarrow \infty$ really quickly as $\alpha \rightarrow \infty$, 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