Normal samples

Josep Fortiana
Facultat de Matemàtiques i Informàtica, UB

2021-03-15

1 Bivariate normal distribution

Univariate normal distribution A r.v. X has a normal (Gaussian) distribution with parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$, $\mathrm{N}(\mu, \sigma^2)$, when it is absolutely contínuous, with probability density function:

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

If $X \sim \mathrm{N}(\mu, \sigma^2)$, then $Z = \dfrac{X-\mu}{\sigma} \sim \mathrm{N}(0,1)$, $X = \mu + \sigma \cdot Z$, and every univariate normal r.v. can be thus obtained.

Bivariate normal distribution with independent marginals If two r.v. , X, Y, have $\mathrm{N}(\mu_x, \sigma_x^2)$ and $\mathrm{N}(\mu_y, \sigma_y^2)$ distributions, respectively, and they are independent, then the vector (X, Y), is absolutely contínuous, with bivariate probability density function:

$$f_{(X,Y)}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left\{-\frac{1}{2}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right]\right\}$$

General bivariate normal distribution A random vector $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ has a bivariate normal distribution if, and only if, there exist two independent standard gaussian r.v. 's Z_1, Z_2, a 2×2 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad \text{and a vector } \; m = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}$$

such that:

$$X = A \cdot Z + m, \quad \text{where } \; Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}.$$

Singular bivariate normal distributions When $\operatorname{rank}(A) = 1$, A is of the form $A = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \cdot (b_1, b_2)$, then

X is singular, having its probability concentrated on a straight line.

Exercise: Compute the equation of this straight line in terms of a_1, a_2, m_1, and m_2.

General nonsingular bivariate normal distribution A random vector (X, Y) has a non singular bivariate normal (Gaussian) distribution if it is absolutely continuous on $\mathbb{R}^2$ and its joint bivariate pdf is of the form:

$$C \exp(-Q/2),$$

where C is a normalizing constant and Q is a positive definite quadratic form:

$$Q = c_{11}(x-\mu_x)^2 + 2\,c_{12}(x-\mu_x)(y-\mu_y) + c_{22}(y-\mu_y)^2.$$

The coefficients in Q can be related to known quantities through straightforward computations.

Firstly we write Q as a sum of squares:

$$Q = c_{22}\left((y-\mu_y) + \frac{c_{12}}{c_{22}}(x-\mu_x)\right)^2 + \left(c_{11} - \frac{c_{12}^2}{c_{22}}\right)(x-\mu_x)^2.$$

Defining:

$$u = (y-\mu_y) + \frac{c_{12}}{c_{22}}(x-\mu_x), \qquad v = (x-\mu_x),$$

we can write:

$$c_{22}\,u^2 + \left(c_{11} - \frac{c_{12}^2}{c_{22}}\right) v^2,$$

showing u and v as two independent centered univariate normal variates, with:

$$\operatorname{var}(u) = 1/c_{22}, \quad \operatorname{var}(v) = c_{22}/\Delta, \quad \text{where } \Delta = c_{11}\,c_{22} - c_{12}^2.$$

Since $v = x - \mu_x$, $\mathrm{E}(x) = \mu_x$ and $\sigma_x^2 \equiv \operatorname{var}(x) = c_{22}/\Delta$. By symmetry, $\mathrm{E}(y) = \mu_y$ and $\sigma_y^2 \equiv \operatorname{var}(y) = c_{11}/\Delta$.

Using the bilinearity of $\operatorname{cov}(\,\cdot\,,\,\cdot\,)$,

$$0 = \operatorname{cov}(u, v) = \operatorname{cov}(x, y) + \frac{c_{12}}{c_{22}}\,\sigma_x^2.$$

From the above equation, $\sigma_x^2/c_{22} = 1/\Delta$, and as a function of $\rho \equiv \operatorname{cov}(x, y)/(\sigma_x\sigma_y)$:

$$-c_{12}/\Delta = \operatorname{cov}(x, y) = \rho\sigma_x\sigma_y.$$

Immediately: $\sigma_x^2\sigma_y^2(1-\rho^2) = 1/\Delta$.

Finally, we have the coefficients of Q as a function of the first two moments:

$$c_{11} = \frac{1}{\sigma_x^2(1-\rho^2)}, \quad c_{22} = \frac{1}{\sigma_y^2(1-\rho^2)}, \quad c_{12} = -\frac{\rho}{\sigma_x\sigma_y(1-\rho^2)},$$

$$Q = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)\right]$$

General bivariate normal pdf in terms of moments An absolutely continuous random vector (X, Y), is bivariate normal with parameters $(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$ if its probability density function is:

$$\begin{aligned} f_{(X,Y)}(x, y) \quad = \quad & \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\times\right. \\ & \left.\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)\right]\right\} \end{aligned}$$

The (univariate) marginals of this vector are univariate normal.

Standardization As a function of the standardized vector (z_x, z_y),

$$z_x = \frac{x - \mu_x}{\sigma_x}, \quad z_y = \frac{y - \mu_y}{\sigma_y},$$

the pdf is:

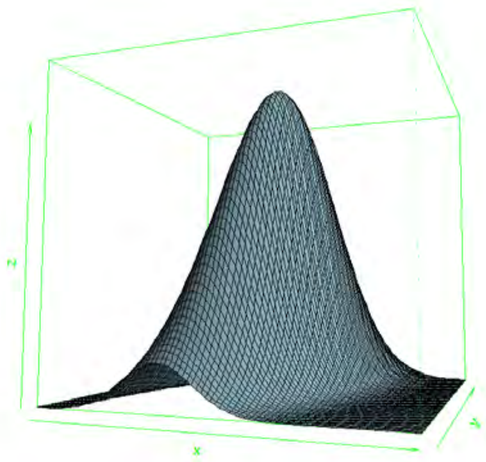
$$f(z_x, z_y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{z_x^2 + z_y^2 - 2\rho z_x z_y}{2(1-\rho^2)}\right\},$$

Contours of Gaussian pdf's The level or contour curves, such that the pdf is constant, are the ellipses:

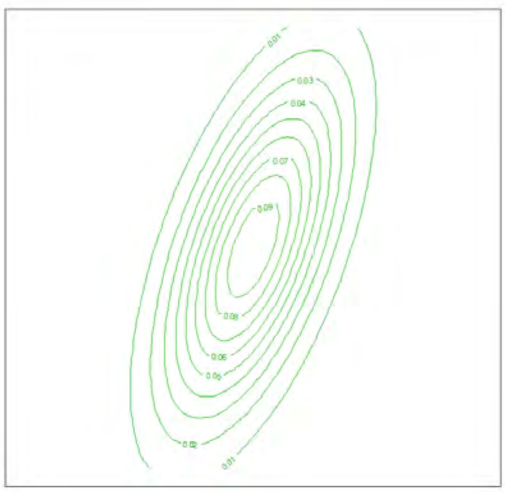
$$Q(x, y) = \text{const.}$$

We could calculate their canonical form, major and minor axes, angle of the principal coordinate system (major and minor axes as coordinate axes) with respect to the usual one, excentricity, etc.

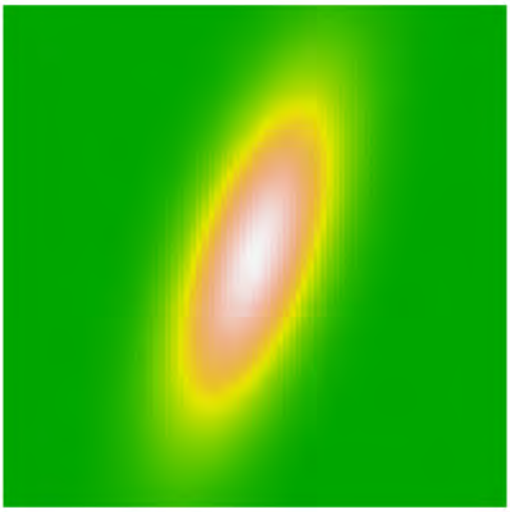
3D plot of a bivariate normal pdf



Contour plot of a bivariate normal pdf



Another version of the contour plot



Conditional pdf Given $x \in \mathbb{R}$ the conditional pdf of $(Y|X=x)$ is:

$$C\ \exp\left\{-\frac{c_{22}}{2}\left((y-\mu_y)+\frac{c_{12}}{c_{22}}(x-\mu_x)\right)^2\right\}$$

$$= \quad C\ \exp\left\{-\frac{1}{2\sigma_y^2(1-\rho^2)}\left((y-\mu_y)+\frac{\rho\sigma_y}{\sigma_x}(x-\mu_x)\right)^2\right\}.$$

The conditional expectation $\mu_{y|x} \equiv \mathrm{E}(Y|X=x)$ is:

$$\mu_{y|x} = \mu_y + \frac{\rho\,\sigma_y}{\sigma_x}\,(x-\mu_x),$$

can also be written:

$$\mu_{y|x} = \beta_0 + \beta_1\,x, \quad \text{where } \beta_1 = \frac{\rho\,\sigma_y}{\sigma_x}, \quad \beta_0 = \mu_y - \beta_1\,\mu_x.$$

and the conditional variance:

$$\sigma_{y|x}^2 = \sigma_y^2\,(1-\rho^2).$$

Moments in matrix form If (X, Y) is a bivariate normal vector with parameters $(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$, then:

$$\mathrm{E}(X, Y) = (\mu_x, \mu_y),$$

and the matrix of variances and covariances (or just *covariance matrix*) is:

$$\mathrm{Var}(X, Y) = \mathbf{\Sigma} = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix},$$

$$\mathrm{cov}(X, Y) = \rho\sigma_x\sigma_y, \qquad \mathrm{cor}(X, Y) = \rho.$$

Inverse of $\mathbf{\Sigma}$

$$\det \mathbf{\Sigma} = \sigma_x^2\,\sigma_y^2\,(1-\rho^2).$$

If $\rho \neq 1$, $\det \mathbf{\Sigma} \neq 0$, and:

$$\mathbf{\Sigma}^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} \dfrac{1}{\sigma_x^2} & -\dfrac{\rho}{\sigma_x\sigma_y} \\ -\dfrac{\rho}{\sigma_x\sigma_y} & \dfrac{1}{\sigma_y^2} \end{pmatrix},$$

Quadratic form in matrix notation The exponent in the pdf,

$$-\frac{1}{2}\,Q(x,y),$$

(if $\rho \neq 1$), is the quadratic form:

$$Q(x,y) = \boldsymbol{u}' \cdot \boldsymbol{\Sigma}^{-1} \cdot \boldsymbol{u},$$

$$\text{where } \boldsymbol{u} = \begin{pmatrix} x-\mu_x \\ y-\mu_y \end{pmatrix}.$$

Pdf as a function of the quadratic form Finally, the pdf in matrix form:

$$f_{(X,Y)}(x,y) = \frac{1}{2\pi\sqrt{\det \boldsymbol{\Sigma}}} \exp\left\{-\frac{1}{2}\,\boldsymbol{u}' \cdot \boldsymbol{\Sigma}^{-1} \cdot \boldsymbol{u}\right\},$$

$$\text{where } \boldsymbol{u} = \begin{pmatrix} x-\mu_x \\ y-\mu_y \end{pmatrix}.$$

This expression is directly generalizable to a p-dimensional Gaussian pdf.

Cautionary remarks

1. There are non-absolutely continuous bivariate normal distributions (they are singular with respect to the natural measure in $\mathbb{R}^2$ and have no pdf).

2. In particular, if F and G are univariate normal, $H_+(F,G)$ and $H_-(F,G)$ are singular distributions, with the whole probability on a straight line –with, respectively, $\rho = +1$ and $\rho = -1$.

3. There are bivariate distributions whose marginals are univariate normal but not bivariate normal themselves.

A non normal bivariate distribution with normal marginals Let $X_1 \sim \mathrm{N}(0,1)$. Define $X_2 = h(X_1)$, where:

$$h(x) = \begin{cases} -x, & \text{if } -1 \leq x \leq 1, \\ x, & \text{otherwise.} \end{cases}$$

H, the joint distribution of (X_1, X_2) satisfies that, by definition, its X_1 marginal is $\mathrm{N}(0,1)$.

We see that:

- The X_2 marginal also is $\mathrm{N}(0,1)$.
- H is not bivariate normal.

We compute the cdf of X_2: For $a \in \mathbb{R}$, $\mathrm{P}\{X_2 \leq a\} = \mathrm{P}\{X_1 \leq a\}$, thus $X_2 \sim \mathrm{N}(0,1)$.

Consider separately the three cases $a < -1$, $-1 \leq a \leq 1$ and $a \geq 1$.

The 1-dimensional subset of $\mathbb{R}^2$ (thus of null measure) has probability > 0,

$$\mathrm{P}\{X_1 - X_2 = 0\} = \mathrm{P}\{|X_1| > 1\} > 0.$$

Hence H is not absolutely continuous on $\mathbb{R}^2$, in particular it is not nonsingular bivariate normal.

Furthermore, it cannot be singular bivariate normal, as no straight line in $\mathbb{R}^2$ has probability 1.

2 Normal data with unknown mean, known variance

Assume $X \sim \mathrm{N}(\theta, \sigma^2)$, where σ is known (fixed, constant), and the prior distribution of θ is also a gaussian: $\theta \sim \mathrm{N}(\mu, \gamma^2)$.

We will see that θ's posterior distribution is also normal, with parameters:

$$
\begin{aligned}
\mathrm{E}(\theta|x) &= \mu_x \stackrel{\text{def}}{=\!=} \frac{\gamma^2}{\sigma^2+\gamma^2}\,x + \frac{\sigma^2}{\sigma^2+\gamma^2}\,\mu \\
\mathrm{var}(\theta|x) &= \tau^2 \stackrel{\text{def}}{=\!=} \frac{\sigma^2\gamma^2}{\sigma^2+\gamma^2} = \frac{1}{\dfrac{1}{\sigma^2}+\dfrac{1}{\gamma^2}}
\end{aligned}
$$

Interpretation of the updated parameters:

- μ_x is a convex combination of μ, the prior expectation, and the observed x.
- Relative weight is inversely proportional to the variances. Precision.
- The *precision* concept also illuminates the meaning of the posterior variance. Precision is *additive*:

$$\frac{1}{\tau^2} = \frac{1}{\sigma^2} + \frac{1}{\gamma^2}$$

Computational details:

The pdf (or likelihood) of x, for a given θ, is:

$$f(x|\theta) = \frac{1}{\sqrt{2\pi}\,\sigma}\,\exp\left\{-\frac{1}{2}\frac{(x-\theta)^2}{\sigma^2}\right\}$$

The prior pdf of θ, for given μ, γ^2, is:

$$h(\theta) = \frac{1}{\sqrt{2\pi}\,\gamma}\,\exp\left\{-\frac{1}{2}\frac{(\theta-\mu)^2}{\gamma^2}\right\}$$

To compute the posterior pdf of θ, for a given x, with Bayes' formula for pdf's, we have to:

- Compute the joint pdf of (x, θ):
$$h(x,\theta) = f(x|\theta)\,h(\theta),$$
- Integrate with respect to θ, to obtain $f(x)$, the marginal pdf of x in the denominator of Bayes' formula:
$$h(\theta|x) = \frac{h(x,\theta)}{f(x)}.$$

Joint distribution of (x, θ):

The exponent in the product $h(x, \theta)$ is:

$$\left\{-\frac{1}{2}\frac{(x-\theta)^2}{\sigma^2}-\frac{1}{2}\frac{(\theta-\mu)^2}{\gamma^2}\right\}$$

$$=-\frac{1}{2}\left\{\frac{\gamma^2(x^2-2x\theta+\theta^2)+\sigma^2(\theta^2-2\mu\theta+\mu^2)}{\sigma^2\gamma^2}\right\}$$

$$=-\frac{1}{2}\left\{\frac{\theta^2(\sigma^2+\gamma^2)-2\theta(\mu\sigma^2+x\gamma^2)+(x^2\gamma^2+\mu^2\sigma^2)}{\sigma^2\gamma^2}\right\}$$

$h(x,\theta)=f(x|\theta)\,h(\theta)$ is a bivariate normal.

However, it is NOT the product of its two marginals!!

The correlation coefficient is:

$$\rho^2(x,\theta)=\frac{\gamma^2}{\gamma^2+\sigma^2}.$$

Obtaining the marginal $f(x)$

Divide both numerator and denominator by $(\sigma^2+\gamma^2)$,

$$=-\frac{1}{2}\left\{\frac{\theta^2-2\theta\mu_x+\dfrac{(x^2\gamma^2+\mu^2\sigma^2)}{\sigma^2+\gamma^2}}{\tau^2}\right\}$$

Completing the square, we find a first summand:

$$-\frac{1}{2}\left\{\frac{\theta^2-2\theta\mu_x+\mu_x^2}{\tau^2}\right\}=-\frac{1}{2}\left\{\frac{(\theta-\mu_x)^2}{\tau^2}\right\},$$

and a second summand not depending on θ which, simplifying, gives:

$$-\frac{1}{2}\left\{\frac{(x-\mu)^2}{\sigma^2+\gamma^2}\right\}.$$

The exponential of the first part is almost a normal pdf for θ. It would require multiplying by the normalization constant $1/\sqrt{2\pi\tau^2}$.

We do this (and compensate the operation multiplying by $\sqrt{2\pi\tau^2}$ the constant already in front of $h(x,\,\theta)$).

Integration of the first part with respect to θ gives 1, thus:

$$f(x)=(2\pi)^{-1/2}\,(\sigma^2+\gamma^2)^{-1/2}\exp\left\{-\frac{1}{2}\left[\frac{(x-\mu)^2}{\sigma^2+\gamma^2}\right]\right\},$$

that is, this marginal pdf is a: $\mathrm{N}(\mu,(\sigma^2+\gamma^2))$.

The x marginal:

$$f(x) = (2\pi)^{-1/2}\,(\sigma^2+\gamma^2)^{-1/2}\exp\left\{-\frac{1}{2}\left[\frac{(x-\mu)^2}{\sigma^2+\gamma^2}\right]\right\},$$

a $\mathrm{N}(\mu, (\sigma^2+\gamma^2))$.

This distribution is called a *prior predictive pdf.*

Average of $f(x|\theta)$ over all possible values of θ, each with its relative weight, according to the prior $h(\theta)$.

Obtaining the posterior pdf of θ, given x:

Dividing $h(x,\theta)$ by $f(x)$, we obtain the posterior pdf:

$$h(\theta|x) = (2\pi)^{-1/2}\,\tau^{-1}\exp\left\{-\frac{1}{2}\frac{(\theta-\mu_x)^2}{\tau^2}\right\},$$

a normal distribution, with expectation:

$$\mu_x = \frac{\sigma^2}{\sigma^2+\gamma^2}\,\mu + \frac{\gamma^2}{\sigma^2+\gamma^2}\,x,$$

and variance:

$$\tau^2 = \frac{\sigma^2\gamma^2}{\sigma^2+\gamma^2}.$$

The *posterior predictive pdf,* $f(\tilde{x}|x)$, of a new observation $\tilde{x}$, given the previously observed value x.

By definition, $f(\tilde{x}|x)$ is the average of $f(\tilde{x}|\theta)$ over all possible values of θ, each with its relative weight, now according to $h(\theta|x)$, the posterior pdf of θ given x.

No new computation is needed. Only comparison with the prior predictive pdf.

Result: the posterior predictive pdf of a new $\tilde{x}$, given x, is a normal distribution:

$$(\tilde{x}|x) \sim \mathrm{N}(\mu_x, \sigma^2+\tau^2), \quad \text{where, as above,}$$

$$\mu_x = \frac{\sigma^2}{\sigma^2+\gamma^2}\,\mu + \frac{\gamma^2}{\sigma^2+\gamma^2}\,x,$$

$$\tau^2 = \frac{\sigma^2\gamma^2}{\sigma^2+\gamma^2}.$$

Case of an n-sample

The often-found case of observing an n-sample, $X_1,\ldots,X_n$ i.i.d.$\sim \mathrm{N}(\theta,\sigma^2)$ can be treated with the formulation above.

Indeed, the observed n-sample, $X_1,\ldots,X_n$, for the purpose of studying θ and according to the *Principle of Sufficiency,* is equivalent to a single observation of $\overline{X} \sim \mathrm{N}(\theta, \sigma^2/n)$.

In particular, we obtain then the posterior parameters of θ:

$$\begin{aligned}
\mathrm{E}(\theta|x) &= \mu_x \stackrel{\text{def}}{=} \frac{\gamma^2}{\sigma^2/n+\gamma^2}\,x + \frac{\sigma^2/n}{\sigma^2/n+\gamma^2}\,\mu \\
\mathrm{var}(\theta|x) &= \tau^2 \stackrel{\text{def}}{=} \frac{\sigma^2\,\gamma^2}{\sigma^2+n\,\gamma^2}
\end{aligned}$$

3 Gamma, chi-squared et cætera

Gamma distribution pdf

The Gamma(α, β) probability distribution with *shape* parameter α and *rate* parameter β is defined by the pdf:

$$f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot x^{\alpha-1}\, e^{-\beta x}, \quad x \geq 0, \quad \alpha, \beta > 0.$$

For $X_1, \ldots, X_n$ independent r.v. , with distributions $X_i \sim \text{Gamma}(\alpha_i, \beta)$, $1 \leq i \leq n$, with the same rate parameter β, the sum:

$$S = \sum_{i=1}^n X_i \sim \text{Gamma}(\sum_{i=1}^n \alpha_i, \beta),$$

is also Gamma-distributed, with the shape parameter equal to the sum of the shape parameters of the summands.

The Exp(β) probability distribution with *rate* parameter β is defined by the pdf:

$$f(x|\beta) = \beta \cdot e^{-\beta x}, \quad x \geq 0, \quad \beta > 0.$$

Clearly $\text{Exp}(\beta) \equiv \text{Gamma}(1, \beta)$.

Additivity: for $X_1, \ldots, X_n$ independent exponential r.v. , with the same parameter β, the sum:

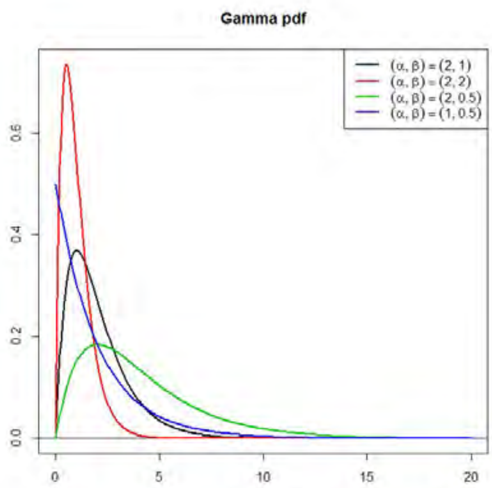
$$S = \sum_{i=1}^n X_i \sim \text{Gamma}(n, \beta).$$

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

$$\text{E}(X) = \frac{\alpha}{\beta}, \quad \text{var}(X) = \frac{\alpha}{\beta^2}.$$

The mode is:

$$\frac{\alpha - 1}{\beta}, \quad \text{for } \alpha > 1.$$



The $\chi^2(k)$ probability distribution

The $\chi^2(k)$ or χ^2_k, the *chi squared distribution* with k *degrees of freedom*, is a Gamma(α, β) with *shape* $\alpha = \frac{k}{2}$ and *rate* $\beta = \frac{1}{2}$. It has the pdf:

$$f(x|k) = \frac{1}{2^{\frac{k}{2}} \cdot \Gamma(\frac{k}{2})} \cdot x^{\frac{k}{2}-1} \cdot e^{-\frac{x}{2}}, \quad x > 0, k > 0.$$

It deserves a special name due to its origin from the normal distribution.

If $X \sim \mathrm{N}(0,1)$ then $Q = X^2 \sim \chi^2(1)$.

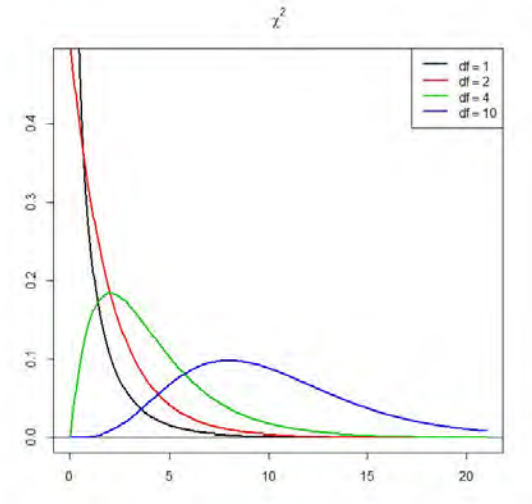
Obviously this is the χ^2 name origin.

More generally, if $X_1, \ldots, X_n$ are i.i.d.$\sim \mathrm{N}(0,1)$, then:

$$Q_n \equiv \sum_{i=1}^n X_i^2 \sim \chi^2(n).$$

A large majority of quantities used as goodness-of-fit/prediction quality measures are $\chi^2(n)$-distributed.

Reason is individual errors tend to be normally distributed and the sum of squared errors is a sensible measure of global error.



$\mathrm{Gamma}(\alpha,\beta)$ **in terms of** $\chi^2(k)$ If $X \sim \mathrm{Gamma}(\alpha,\beta)$, the new r.v. defined as:

$$Z = 2\,\beta\,X, \quad \frac{Z}{2} = \beta\,X, \quad X = \frac{1}{2\,\beta}\,Z,$$

has pdf:

$$f_Z(z) = \frac{1}{2^\alpha \cdot \Gamma(\alpha)} \cdot z^{\alpha-1} \cdot e^{-\frac{z}{2}}, \quad z > 0,$$

i.e., a χ^2, with $k = 2\,\alpha$ degrees of freedom. Thus, a Gamma distribution may be considered as a scaled χ^2.

The inverse gamma distribution

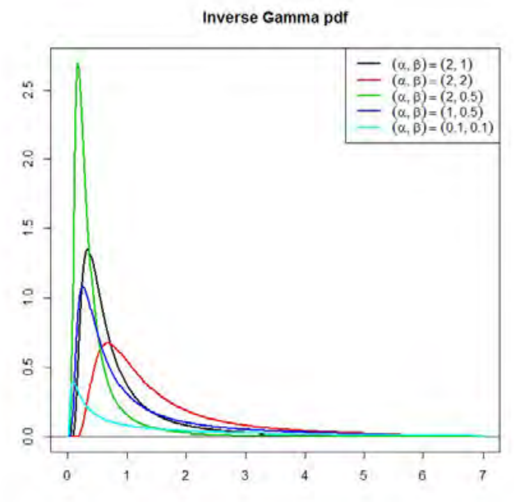
When $X \sim \mathrm{Gamma}(\alpha,\beta)$, then $Y = \dfrac{1}{X}$, by definition,

has an *inverse gamma distribution* $\mathrm{IG}(\alpha,\beta)$.

Its pdf is:

$$f_Y(y|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \frac{1}{y^{\alpha+1}} \cdot e^{-\frac{\beta}{y}}, \quad y > 0,\ \beta > 0.$$

Warning: parameter β in the $\text{IG}(\alpha, \beta)$ is called the scale parameter, i.e., the converse nomenclature of that in the $\text{Gamma}(\alpha, \beta)$ distribution.



For $Y \sim \text{IG}(\alpha, \beta)$,

$$\text{E}(Y) = \frac{\beta}{\alpha - 1}, \qquad \text{for } \alpha > 1,$$

$$\text{var}(Y) = \frac{\beta^2}{(\alpha - 1)^2 (\alpha - 2)}, \qquad \text{for } \alpha > 2.$$

The mode is:

$$\text{Mode}(Y) = \frac{\beta}{\alpha + 1}.$$

Inverse chi squared distribution

The inverse chi squared distribution with k degrees of freedom, $\text{Inv-}\chi^2(k)$, is an $\text{IG}(\alpha = \frac{k}{2}, \beta = \frac{1}{2})$. Its pdf is:

$$f(z) = \frac{2^{-k/2}}{\Gamma(k/2)} z^{-k/2-1} e^{-1/(2z)}, \quad z > 0.$$

As in the case of the $\text{Gamma}(\alpha, \beta)$ with the $\chi^2(k)$, some authors tend to write an $\text{IG}(\alpha, \beta)$ as a scaled $\text{Inv-}\chi^2$ distribution.

Expectation, variance, mode of $\text{Inv-}\chi^2(k)$

For $Z \sim \text{Inv-}\chi^2(k)$,

$$\text{E}(Z) = \frac{1}{k - 2}, \qquad \text{for } k > 2,$$

$$\text{var}(Z) = \frac{2}{(k - 2)^2 (k - 4)}, \qquad \text{for } k > 4.$$

The mode is:

$$\text{Mode}(Z) = \frac{1}{k+2}.$$

4 Known mean, unknown variance

Likelihood: Given n i.i.d. normal observations, $\boldsymbol{x} = (x_1, \ldots, x_n)$, whose variance is unknown but whose expectation is known, assumed 0, the likelihood is:

$$f(\boldsymbol{x}\,|\psi) = (2\pi)^{-n/2} \cdot \psi^{n/2} \cdot \exp\left\{-\frac{n\,s^2}{2} \cdot \psi\right\},$$

where $\psi = \dfrac{1}{\sigma^2}$ is the precision parameter, and

$s^2 = \dfrac{1}{n} \sum_{i=1}^{n} x_i^2$ is the empirical variance.

Conjugate prior for the precision parameter: In terms of the precision ψ the conjugate pdf is Gamma(α, β).

$$h(\psi|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \psi^{\alpha-1} \cdot \exp\{-\beta\,\psi\}.$$

This conjugate distribution, in terms of the variance, is an IG(α, β).

Joint pdf Multiplying likelihood times prior, we obtain the joint pdf: $h(\boldsymbol{x}\,, \psi) =$

$$(2\pi)^{-n/2} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \psi^{(\frac{n}{2}+\alpha-1)} \cdot \exp\left\{-\left(\frac{n\,s^2}{2} + \beta\right) \cdot \psi\right\}.$$

Define:

$$\begin{cases} \widetilde{\alpha} &= \alpha + \dfrac{n}{2}, \\[2ex] \widetilde{\beta} &= \beta + \dfrac{n\,s^2}{2}. \end{cases}$$

Multiply and divide $h(\boldsymbol{x}\,, \psi)$ by $\dfrac{\widetilde{\beta}^{\widetilde{\alpha}}}{\Gamma(\widetilde{\alpha})}$.

Realize the second half is a Gamma$(\widetilde{\alpha}, \widetilde{\beta})$ pdf, which integrates to 1.

Then, the remaining expression is the marginal of $\boldsymbol{x}$:

$$f(\boldsymbol{x}\,) = (2\pi)^{-n/2} \cdot \frac{\Gamma\left(\alpha + \frac{n}{2}\right)}{\Gamma(\alpha)} \cdot \frac{\beta^\alpha}{\left(\beta + n\,\frac{s^2}{2}\right)^{(\alpha+\frac{n}{2})}}.$$

Marginal pdf - Prior predictive pdf $\boldsymbol{x}$ appears only through $s^2 = \dfrac{1}{n} \sum_{i=1}^{n} x_i^2$ *(Sufficiency).*

$$\text{Define: } k = 2\,\alpha + n - 1, \quad t^2 = k \cdot \frac{n\,s^2}{2\,\beta}.$$

The marginal pdf, in terms of t^2, is proportional to:

$$f(t) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\,\pi} \cdot \Gamma\left(\frac{k}{2}\right)} \cdot \left(1 + \frac{t^2}{k}\right)^{-\frac{k+1}{2}}, \quad -\infty < t < \infty,$$

a Student's $t(k)$ pdf.

Posterior pdf of $(\psi|\boldsymbol{x})$ From Bayes' rule we see that:

$$(\psi|\boldsymbol{x}) \sim \text{Gamma}(\widetilde{\alpha}, \widetilde{\beta}),$$

where:

$$\begin{cases} \widetilde{\alpha} &= \alpha + \dfrac{n}{2}, \\ \widetilde{\beta} &= \beta + \dfrac{n\, s^2}{2}. \end{cases}$$

5 Normal data with both parameters unknown

The likelihood function for n i.i.d.$\sim$ N(μ, ψ) normal observations, $\boldsymbol{x} = (x_1, \ldots, x_n)$, $\psi = 1/\sigma^2$, is:

$$f(\boldsymbol{x}\,|\mu, \psi) \propto \psi^{n/2} \cdot \exp\left\{-\frac{\psi}{2} \textstyle\sum_{i=1}^{n}(x_i - \mu)^2\right\},$$

We assume now that both parameters (μ, ψ) are unknown, hence we must provide prior pdf's for both of them.

Joint prior pdf This is the new feature when there is more than one prior parameter: we need a joint prior pdf for (μ, ψ).

Indeed we could try to assume that μ and ψ are independent, by posing a prior pdf:

$$h(\mu, \psi) = h_1(\mu) \cdot h_2(\psi),$$

but then we would not obtain a conjugate prior.

We propose:

$$\psi \sim \text{Gamma}(\alpha, \beta),$$

$$\mu|\psi \sim \text{N}(\mu_0, n_0\, \psi).$$

n_0 is a scaling factor parameter analogous to the number of observations in a virtual "prior sample".

That is, the variance of the mean of n_0 observations, each with variance σ^2, is σ^2/n_0, corresponding to the precision $n_0\, \psi$, where $\psi = 1/\sigma^2$.

Posterior for μ**, given** x **and** ψ We already did this computation:

$$\mu|(\boldsymbol{x}, \psi) \sim \text{N}(\mu_x, \psi_x),$$

where:

$$\mu_x = \frac{n}{n + n_0}\, \bar{x} + \frac{n_0}{n + n_0}\, \mu_0,$$

$$\psi_x = (n + n_0) \cdot \psi.$$

Posterior for ψ, given x

$$(\psi|\boldsymbol{x}) \sim \text{Gamma}(\widetilde{\alpha}, \widetilde{\beta}),$$

where:

$$\begin{cases} \widetilde{\alpha} & = & \alpha + \dfrac{n}{2}, \\ \\ \widetilde{\beta} & = & \beta + \dfrac{n\, s^2}{2} + \dfrac{n \cdot n_0}{2\,(n + n_0)}\,(\bar{x} - \mu_0)^2. \end{cases}$$

Here $s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/n$.
