Summary of Normal Distribution[1]

Normal (Gaussian) random variables

1. We write $X \sim \mathcal{N}(\mu, \sigma^2)$ and say that the random variable X is normal, or Gaussian, with mean μ and variance σ^2, if X is an absolutely continuous random variable with pdf $\dfrac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$, $x \in \mathbb{R}$. In this case,

$$Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$$

is called `standard normal`.

2. If $Z \sim \mathcal{N}(0,1)$, and $a > 0$, then
   - $P(-a < Z < 0) = P(0 < Z < a)$;
   - $P(Z > a) = P(Z < -a) = 0.5 - P(0 < Z < a)$;
   - $P(|Z| > a) = 2P(Z > a)$;
   - $P(Z < a) = P(Z > -a) = 0.5 + P(0 < Z < a)$.

Note: $P(Z < a) > 0.5$ if and only if $a > 0$; $P(Z > b) > 0.5$ if an only if $b < 0$. For example,
   - $P(Z < 1.1) = 0.5 + P(0 < Z < 1.1) = 0.8643$;
   - If you know that $P(Z > c) = 0.6179$, then $c < 0$ and $P(0 < Z < |c|) = 0.1179$, which means that $|c| = 0.3$ and $c = -0.3$.

Drawing a picture of the "Bell Curve" is very helpful

3. If $Y_1, \ldots, Y_n$ are independent so that $Y_k \sim \mathcal{N}(\mu_k, \sigma_k^2)$ and $a_1, \ldots, a_n$ are real numbers, then $a_1Y_1 + \ldots + a_nY_n \sim \mathcal{N}(\mu, \sigma^2)$, where

$$\mu = a_1\mu_1 + \ldots + a_n\mu_n, \quad \sigma^2 = a_1^2\sigma_1^2 + \ldots + a_n^2\sigma_n^2.$$

In particular,
   - If $Y_k,\ k = 1, \ldots, n$, are iid[2] $\mathcal{N}(\mu, \sigma^2)$, then $Y_1 + \ldots + Y_n \sim \mathcal{N}(n\mu, n\sigma^2)$.
   - If $Y_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ are independent, then $Y_1 - Y_2 \sim \mathcal{N}(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$.
   - In general, if $Y_k,\ k = 1, \ldots, n$, are independent $\mathcal{N}(\mu_k, \sigma_k^2)$ and $a_k,\ k = 1, \ldots, n$, are real numbers, then $\sum_{k=1}^n a_kY_k$ is $\mathcal{N}\left(\sum_{k=1}^n a_k\mu_k, \sum_{k=1}^n a_k^2\sigma_k^2\right)$.

The Central Limit Theorem (CLT)

1. **Basic result:** if $X_1, \ldots, X_n$ are iid with mean μ and standard deviation σ, and $n > 30$, then $X_1 + \ldots + X_n$ is approximately normal with mean $n\mu$ and standard deviation $\sqrt{n}\sigma$, while the sample mean $\bar{X}_n = (X_1 + \ldots + X_n)/n$ is approximately normal with mean μ and standard deviation $\sigma/\sqrt{n}$. Equivalently,

$$\lim_{n\to\infty} \frac{X_1 + \ldots + X_n - n\mu}{\sigma\sqrt{n}} = \mathcal{N}(0,1) \quad \text{(in distribution)}.$$

2. **CLT for Binomial distribution:** if $np(1-p) > 5$, then $\mathcal{B}(n,p) \approx \mathcal{N}(np, np(1-p))$.
In the problems:
   (1) Identify the "success" event.
   (2) Compute the probability of success p.
   (3) Check that $np(1-p) > 5$.
   (4) Use continuity correction by enlarging *closed* intervals. For example, $P(X > m) = P(X \geq m+1) = P(X > m + 0.5)$.
   (5) Normalize to get standard normal: $\dfrac{X - np}{\sqrt{np(1-p)}} \approx \mathcal{N}(0,1)$.

3. **CLT for general distributions:**
   (1) Compute the expected value and standard deviation for the distribution. If the distribution is continuous, you *might* need integration.

---
[1] Sergey Lototsky, USC
[2] independent and identically distributed

(2) Check whether the question is asking for the sum, or for the sample mean, or for something else, and then use appropriate normalization.
(3) If the distribution is discrete, use continuity correction, by enlarging *closed* intervals.

Normal (Gaussian) Vectors

Below, $\mathfrak{i} = \sqrt{-1}$, $(\cdot,\cdot)$ is inner product in Euclidean space $\mathbb{R}^n$; C^{-1} means inverse of the matrix C; C^T means the transpose of C. Vectors written in bold face and are thought of as matrices with one column.

The following **three definitions** of a Gaussian vector $\boldsymbol{X} = (X_1, \ldots, X_n)$ are equivalent:

(1) $\mathbb{E}e^{\mathfrak{i}(\boldsymbol{X},\boldsymbol{\lambda})} = e^{\mathfrak{i}(\boldsymbol{\lambda},\boldsymbol{\mu})-(1/2)(C\boldsymbol{\lambda},\boldsymbol{\lambda})}$, $\boldsymbol{\lambda} \in \mathbb{R}^n$, for some vector $\boldsymbol{\mu}$ and a symmetric non-negative definite matrix C [that is, $C = C^T$ and $(C\boldsymbol{a},\boldsymbol{a}) \geq 0$ for all $\boldsymbol{a} \in \mathbb{R}^n$]; with this characterization, $\boldsymbol{\mu} = \mathbb{E}\boldsymbol{X}$ and $C = C_{XX}$ is the covariance matrix of X:

$$C_{XX} = \mathbb{E}\Big((\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^T\Big).$$

THE MAIN CONSEQUENCE OF THIS DEFINITION: *if the entry in column i and row j of the matrix C_{XX} is zero [that is, the random variables X_i and X_j are uncorrelated], then the random variables X_i and X_j are independent.* Also, if the matrix C_{XX} is invertible, with inverse C^{-1}, then, and only then, the vector $\boldsymbol{X}$ has a density (pdf) in $\mathbb{R}^n$:

$$f_X(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(C_{XX})}}\, e^{-\big(\boldsymbol{x}-\boldsymbol{\mu},C^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\big)/2}.$$

(2) $(\boldsymbol{a},\boldsymbol{X}) = \sum_{k=1}^{n} a_k X_k$ is a Gaussian random variable for every *fixed* $\boldsymbol{a} \in \mathbb{R}^n$. THE MAIN PURPOSE OF THIS DEFINITION is *extension to infinite dimensions.*
(3) $\boldsymbol{X} = \boldsymbol{\mu} + A\boldsymbol{Z}$, where $\boldsymbol{Z}$ is a vector with iid standard normal [mean zero variance one] components and A is a square matrix. In this case, $C_{XX} = AA^T$. THE MAIN CONSEQUENCE OF THIS DEFINITION is *different representations of the normal vector.*

Representations of the normal vector $\boldsymbol{X}$, with mean $\boldsymbol{\mu}$ and covariance C_{XX}, using a vector $\boldsymbol{Z}$ with iid standard normal [mean zero variance one] components.

(1) The `Karhunen-Loève`, or `KL`, expansion (finite-dimensional version) of $\boldsymbol{X}$ is

$$\boldsymbol{X} = \boldsymbol{\mu} + QR^{1/2}\boldsymbol{Z},$$

where the *orthogonal* matrix Q and the non-negative diagonal matrix R satisfy $QRQ^T = C_{XX}$.
(2) The `canonical` representation of vector $\boldsymbol{X}$ is

$$\boldsymbol{X} = \boldsymbol{\mu} + L\boldsymbol{Z},$$

where L is the lower-triangular matrix in the `Cholesky decomposition` of C_{XX}: $LL^T = C_{XX}$. If C_{XX} is not invertible, then additional conditions are imposed to ensure uniqueness of L.

The multi-dimensional Normal Correlation Theorem (NCT). Let $\boldsymbol{X}$ be a Gaussian vector in $\mathbb{R}^n$, let $\boldsymbol{Y}$ be a Gaussian vector in $\mathbb{R}^m$. Assume that the combined vector $\boldsymbol{X},\boldsymbol{Y}$ is Gaussian in $\mathbb{R}^{m+n}$ and the covariance matrix C_{YY} of Y is invertible. Then

$$\mathbb{E}(\boldsymbol{X}|\boldsymbol{Y}) = \mathbb{E}\boldsymbol{X} + C_{XY}C_{YY}^{-1}(\boldsymbol{Y}-\mathbb{E}\boldsymbol{Y}),\ \mathbb{E}\Big(\boldsymbol{X}-\mathbb{E}(\boldsymbol{X}|\boldsymbol{Y})\Big)\Big(\boldsymbol{X}-\mathbb{E}(\boldsymbol{X}|\boldsymbol{Y})\Big)^T = C_{XX} - C_{XY}C_{YY}^{-1}C_{YX}.$$

Note that $C_{YX} = C_{XY}^T$.

For the proof, start by finding a matrix A such that the vector $\boldsymbol{X} - \mathbb{E}\boldsymbol{X} - A(\boldsymbol{Y} - \mathbb{E}\boldsymbol{Y})$ and the vector $\boldsymbol{Y} - \mathbb{E}\boldsymbol{Y}$ are uncorrelated. The result: $A = C_{XY}C_{YY}^{-1}$.

If the matrix C_{YY} is not invertible, then the result still holds with the `generalized` or `Moore-Penrose` inverse C_{YY}^{+} of C_{YY}.

If $m = n = 1$ and $\rho = \dfrac{\mathbb{E}(XY) - \mu_X\mu_Y}{\sigma_X\sigma_Y}$ is the correlation coefficient, then the conditional expectation is the equation of the regression line, with X as a function of Y:

$$\mathbb{E}(X|Y) = \mu_X + \rho\frac{\sigma_X}{\sigma_Y}\big(Y - \mu_Y\big).$$

For the conditional variance, $\mathbb{E}\Big(X - \mathbb{E}(X|Y)\Big)\Big(X - \mathbb{E}(X|Y)\Big) = \sigma_X^2(1-\rho^2)$. In particular, we have

$$Y = \mu_Y + \sigma_Y Z_1,\ X = \mu_X + \rho\sigma_X Z_1 + \sigma_X\sqrt{1-\rho^2}\, Z_2,$$

where Z_1 and Z_1 are iid standard normal (and the equalities are in distribution).

Normal Approximation: Advanced topics

The delta method. If

$$\lim_{n\to+\infty}\sqrt{n}\big(Z_n - \mu\big) = \mathcal{N}(0,\sigma^2), \quad \text{in distribution},$$

and $f = f(x)$ is a continuously differentiable function with $f'(\mu) \neq 0$, then, using Taylor expansion,

$$\lim_{n\to+\infty}\sqrt{n}\big(f(Z_n) - f(\mu)\big) = \mathcal{N}\big(0,\sigma^2|f'(\mu)|^2\big), \quad \text{in distribution}.$$

The result extends to higher dimensions.

Normal approximation of common distributions

(1) For fixed p, Binomial $\mathcal{B}(n,p)$, being a sum of iid $\mathcal{B}(1,p)$ is approximately $\mathcal{N}\big(np, np(1-p)\big)$.

(2) For fixed $\lambda > 0$, Poisson $\mathcal{P}(n\lambda)$ with mean $n\lambda$, being a sum of iid $\mathcal{P}(\lambda)$, is approximately $\mathcal{N}\big(n\lambda, n\lambda\big)$. More generally,

$$\lim_{\gamma\to+\infty}\frac{\mathcal{P}(\gamma)-\gamma}{\sqrt{\gamma}} = \mathcal{N}(0,1), \quad \text{in distribution}.$$

(3) For fixed $a > 0$ and $b > 0$, the Gamma distribution $\mathrm{Gamma}(na, 1/b)$ with shape parameter na and mean value nab, being a sum of iid $\mathrm{Gamma}(a, 1/b)$, is $\mathcal{N}\big(nab, nab^2\big)$

(4) For fixed $a > 0$ and $b > 0$, the Beta distribution $\mathrm{Beta}(na, nb)$ is approximately

$$\mathcal{N}\left(\frac{a}{a+b}, \frac{ab}{n(a+b)^3}\right).$$

One way to see it is to write

$$\mathrm{Beta}(na, nb) = \frac{\mathrm{Gamma}(na,1)}{\mathrm{Gamma}(na,1)+\mathrm{Gamma}(nb,1)},$$

with independent $\mathrm{Gamma}(na,1)$ and $\mathrm{Gamma}(nb,1)$, and then apply the two-dimensional delta method.

Gaussian Processes and Fields

Definition. Given a set $\mathbb{T}$, a collection of random variables $\boldsymbol{X} = \{X(t),\ t\in\mathbb{T}\}$, is called a `Gaussian process` if, for every finite collection $\{t_1,\dots,t_n\}\subset\mathbb{T}$, the random vector $(X(t_1),\dots,X(t_n))$ is Gaussian.

The term "process" is generic. By convention, if $\mathbb{T}\subseteq\mathbb{R}$, then $\boldsymbol{X}$ is the (proper) process, and can also be called a sequence if $\mathbb{T}$ is countable. If $\mathbb{T}\subseteq\mathbb{R}^n$, $n>1$, then $\boldsymbol{X}$ is the field.

The distribution of $\boldsymbol{X}$, as a random object with values in $\mathbb{R}^{\mathbb{T}}$, is defined by

$$\mathbf{P}_X(A) = \mathbb{P}(\boldsymbol{X}\in A),\ A\in\mathcal{B}(\mathbb{R}^{\mathbb{T}}),$$

and is completely determined by two functions: mean value $\mu(t) = \mathbb{E}X(t),\ t\in\mathbb{T}$, and covariance $R(t,s) = \mathbb{E}\big(X(t)X(s)\big) - \mu(t)\mu(s),\ t,s\in\mathbb{T}$. The function R is necessarily `non-negative definite`: for all finite collections $\{t_1,\dots,t_n\}\subset\mathbb{T}$, $\{a_1,\dots,a_n\}\subset\mathbb{R}$,

$$\sum_{k,m=1}^{n} a_k a_m R(t_k,t_m) \geq 0.$$

The reason is simple: $\sum_{k,m=1}^{n} a_k a_m R(t_k,t_m) = \mathbb{E}\Big(\sum_{k=1}^{n} a_k\big(X(t_k)-\mu(t_k)\big)\Big)^2$.

Kolmogorov's Continuity Criterion (Gaussian case, informal statement) If $\mathbb{T} \subseteq \mathbb{R}^n$, $\mathbb{E}X(t)=0$, and $\mathbb{E}\big(X(t)-X(s)\big)^2 \le C|t-s|^{\delta}$ for some $C,\delta>0$, then the function $t \mapsto X(t)$ is continuous (in fact, Hölder continuous of every order less than $\delta/2$.)

A list of zero mean Gaussian processes

Process	Covariance function
Brownian motion (BM)	$\min(t,s),\ t,s \ge 0$
Brownian bridge	$\min(t,s)-ts, \quad t,s \in [0,1]$
Stationary OU	$e^{-\lvert t-s\rvert},\ t,s \in \mathbb{R}$
fractional Brownian motion (fBM)	$\frac{1}{2}\big(\lvert t\rvert^{2H}+\lvert s\rvert^{2H}-\lvert t-s\rvert^{2H}\big), \quad t,s\in\mathbb{R},\ H\in(0,1)$
sub-fractional Brownian motion	$t^{2H}+s^{2H}-\frac{1}{2}\big((t+s)^{2H}+\lvert t-s\rvert^{2H}\big), \quad t,s\ge 0,\ H\in(0,1)$
bi-fractional Brownian motion	$\frac{1}{2^k}\big((t^{2H}+s^{2H})^k-\lvert t-s\rvert^{2Hk}\big),\ t,s\ge 0, H\in(0,1), 0<k\le 2, Hk\le 1$
Brownian sheet	$\prod_{k=1}^{n}\min(x_k,y_k),\ x_k,y_k\ge 0$
Lévy's Brownian motion	$\frac{1}{2}\big(\lvert\boldsymbol{x}\rvert+\lvert\boldsymbol{y}\rvert-\lvert\boldsymbol{x}-\boldsymbol{y}\rvert\big),\ \boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n$
Gaussian free field (GFF)	Green's function of the Laplacian

"General ideas for generalizations" Given a continuous Gaussian process $X=X(t),\ t\ge 0$, with mean zero, $X(0)=0$, and covariance function $R=R(t,s)$, one can construct

- the X-`bridge` $X_b=X_b(t),\ t\in[0,T]$ from $(0,0)$ to $(T,0)$ by defining
$$X_b(t)=X(t)-\frac{R(T,t)}{R(T,T)}X(T),$$
which, by NCT, is conditioning X to hit 0 at time T.
- The X-`OU process` $Y=Y(t)$, $t\ge 0$, as the solution of
$$dY(t)=aY(t)dt+dX(t),\ t\ge 0,\ a\in\mathbb{R}.$$
Note that, solving the equation and integrating by parts,
$$Y(t)=Y(0)e^{at}+\int_0^t e^{a(t-s)}dX(s)=Y(0)e^{at}+X(t)+a\int_0^t X(s)e^{a(t-s)}ds.$$
- The X-`sheet`, as a Gaussian random field with covariance function $\prod_{k=1}^{n} R(x_k,y_k)$.
- The `Volterra` X-`process` with a suitable kernel $K=K(t,s),\ 0\le s\le t$, by $\int_0^t K(t,s)\,dX(s)$.