7. Normal Distribution

Spring 2021

Matthew Blackwell

Gov 2002 (Harvard)

Where are we? Where are we going?

- Learning about r.v. distributions in all combinations
    - Discrete vs. continuous and joint vs. marginal
- Soon: inference (or learning about parameters of these distributions)
- But first: the normal distribution and its cousins.
    - Why: massively important to inference due to the central limit theorem.

Standard normal distribution

Definition

A continuous r.v. Z follows a **standard normal distribution** if its p.d.f. φ is given as

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \qquad -\infty < z < \infty,$$

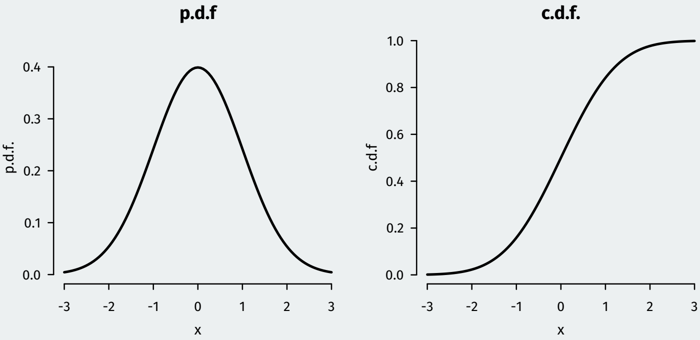
and we write this $Z \sim \mathcal{N}(0,1)$

- Not immediately obvious, but tricky calculus will show $\int_{-\infty}^{\infty} \varphi(z) = 1$.
- Normal c.d.f. has no closed form solution, so written as:

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

- Standard normal is mean zero, variance 1: $\mathbb{E}[Z] = 0, \mathbb{V}[Z] = 1$.

The normal distribution



- Deeply symmetric:
    - p.d.f. is symmetric: $\varphi(z) = \varphi(-z)$
    - Tail areas are symmetric $\Phi(z) = 1 - \Phi(-z)$
    - Z and $-Z$ are both $\mathcal{N}(0, 1)$

General normal distribution

Defintion

If $Z \sim \mathcal{N}(0,1)$ then

$$X = \mu + \sigma Z$$

follows the normal distribution with mean μ and variance σ^2, written $X \sim \mathcal{N}(\mu, \sigma^2)$.

- We can move back to a standard normal through **standardization**:

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

- c.d.f.: $\Phi((x-\mu)/\sigma)$
- p.d.f.:

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

Properties of normals and sums

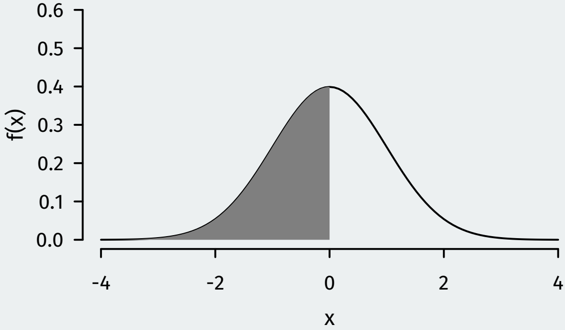
- If $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ and $X_1 \perp\!\!\!\perp X_2$,

$$X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

- **Cramer's theorem:** if $X_1 \perp\!\!\!\perp X_2$ and $X_1 + X_2$ is normal, then X_1 and X_2 are normal.

Using pnorm

- `pnorm()` evaluates the c.d.f. of the normal:

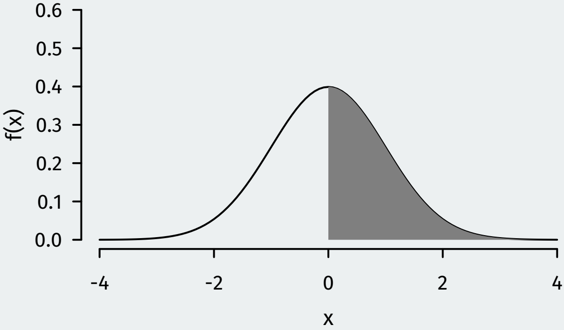



```r
pnorm(q = 0, mean = 0, sd = 1)
```

```
## [1] 0.5
```

Using pnorm

- `pnorm()` evaluates the c.d.f. of the normal:

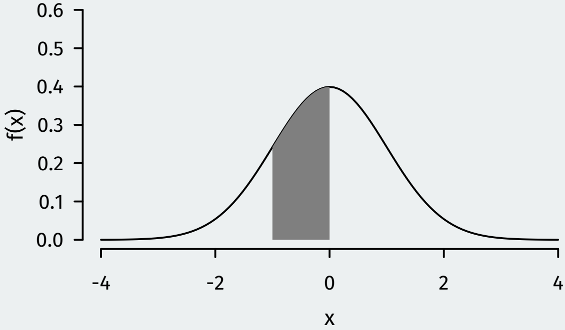



```
pnorm(q = 0, mean = 0, sd = 1, lower.tail = FALSE)
```

```
## [1] 0.5
```

Using pnorm

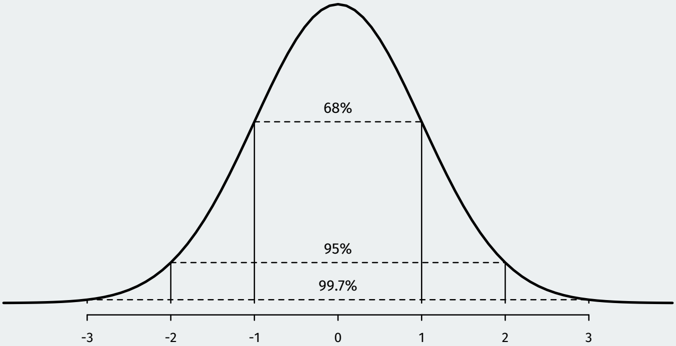
- `pnorm( )` evaluates the c.d.f. of the normal:




```r
pnorm(q = 0, mean = 0, sd = 1) - pnorm(q = -1, mean = 0, sd = 1)
```

```
## [1] 0.341
```

Empirical Rule for the Normal Distribution



- If $Z \sim \mathcal{N}(0, 1)$, then the following are roughly true:
    - Roughly 68% of the distribution of Z is between -1 and 1.
    - Roughly 95% of the distribution of Z is between -2 and 2.
    - Roughly 99.7% of the distribution of Z is between -3 and 3.

Chi-square distribution

Definition

Let $V = Z_1^2 + \cdots + Z_n^2$ where $Z_1, Z_2, \ldots, Z_n$ are i.i.d. $\mathcal{N}(0,1)$. Then V follows the **Chi-square distribution** with n degrees of freedom, written $V \sim \chi_n^2$

- Why do we care? **Sample variance** of normal r.v.s $X_1, \ldots, X_n$ i.i.d. $N(\mu, \sigma^2)$:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \qquad \frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$$

Student t distribution

Definition

If $Z \sim \mathcal{N}(0,1)$ and $V \sim \chi^2_n$ with $Z \perp\!\!\!\perp V$, then

$$T = \frac{Z}{\sqrt{V/n}},$$

follows the **student-t distribution** with n degrees of freedom, written $T \sim t_n$.

- Important result for the **normal model**: if $X_1, \dots, X_n$ are i.i.d. $\mathcal{N}(\mu, \sigma^2)$:

$$T = \frac{\overline{X}_n - \mu}{\sqrt{s^2/n}} \sim t_{n-1}$$

- Properties of the t distribution:
    - Symmetric and mean-zero like the standard normal.
    - Fatter tails than the normal.
    - Converges to $\mathcal{N}(0,1)$ as $n \to \infty$

Multivariate random vectors

- Can group r.v.s into **random vectors** $\mathbf{X} = (X_1, \ldots, X_k)'$
  - $\mathbf{X}$ is a function from the sample space to $\mathbb{R}^k$
  - $\mathbf{x}$ is now a length-k vector and potential value of $\mathbf{X}$
  - Generalizes all ideas from 2 variables to k
- Joint distribution function: $F(\mathbf{x}) = \mathbb{P}(\mathbf{X} \leq \mathbf{x}) = \mathbb{P}(X_1 \leq x_1, \ldots, X_k \leq x_k)$.
  - Discrete: joint p.m.f. $\mathbb{P}(\mathbf{X} = \mathbf{x})$.
  - Continuous: joint p.d.f.

$$f(\mathbf{x}) = \frac{\partial^m}{\partial x_1 \cdots \partial x_k} F(\mathbf{x})$$

- Expectation of a random vector is just the vector of expectations:

$$\mathbb{E}[\mathbf{X}] = (\mathbb{E}[X_1], \mathbb{E}[X_2], \ldots, \mathbb{E}[X_k])'$$

Covariance matrices

- Covariance matrix generalizes (co)variance to this setting:

$$\mathbb{V}[\mathbf{X}] = \mathbb{E}\left[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])'\right]$$

- We usually write $\mathbb{V}[\mathbf{X}] = \mathbf{\Sigma}$ and it is a $k \times k$ **symmetric** matrix:

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \cdots & \sigma_k^2 \end{pmatrix}$$

where, $\sigma_j^2 = \mathbb{V}[X_j]$ and $\sigma_{ij} = \text{Cov}(X_i, X_j)$.

- Symmetric ($\mathbf{\Sigma} = \mathbf{\Sigma}'$) because $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i)$.

Multivariate standard normal distribution

- Let $\mathbf{Z} = Z_1, Z_2, \dots, Z_k$ be i.i.d. $\mathcal{N}(0, 1)$. What is their joint distribution?
- For vector of values $\mathbf{z} = (z_1, z_2, \dots, z_k)^T$

$$f(\mathbf{z}) = \frac{1}{(2\pi)^{k/2}} \exp\left(-\frac{\mathbf{z}'\mathbf{z}}{2}\right)$$

- Easy to see the mean/variance: $\mathbb{E}[\mathbf{Z}] = 0$ and $\mathbb{V}[\mathbf{Z}] = \mathbf{I}_k$.
    - $\mathbf{I}_k$ is the k by k identity matrix because $\mathbb{V}[Z_j] = 1$ and $\text{Cov}(Z_i, Z_j) = 0$.

Linear transformations of random vectors

Theorem

If $\mathbf{X} \in \mathbb{R}^k$ with $k \times 1$ expectation $\boldsymbol{\mu}$ and $k \times k$ covariance matrix $\boldsymbol{\Sigma}$, and $\mathbf{A}$ is a $q \times k$ matrix, then $\mathbf{AX}$ is a random vector with mean $\mathbf{A}\boldsymbol{\mu}$ and covariance matrix $\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'$.

- Let $\mathbf{Z} \sim \mathcal{N}(0, \mathbf{I}_k)$ and $\mathbf{X} = \boldsymbol{\mu} + \mathbf{BZ}$, where $\mathbf{B}$ is $q \times k$ then $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
  - $\boldsymbol{\mu}$: $q \times 1$ mean vector $\mathbb{E}[\mathbf{X}] = \boldsymbol{\mu}$
  - $\boldsymbol{\Sigma} = \mathbf{BB}'$: $q \times q$ covariance matrix $\mathbb{V}[\mathbf{X}] = \boldsymbol{\Sigma}$.
- More generally, if $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ then $\mathbf{Y} = \mathbf{a} + \mathbf{BX} \sim \mathcal{N}(\mathbf{a} + \mathbf{B}\boldsymbol{\mu}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}')$

Properties of the multivariate normal

- If (X_1, X_2, X_3) are MVN, then (X_1, X_2) is also MVN.
- If (X, Y) are multivariate normal with $\text{Cov}(X, Y) = 0$, then X and Y are independent.