Normal distribution

David J. H. Shih

Intended learning outcomes

- Derive the normal distribution
- Explain and apply the Chebychev's inequality, laws of large numbers, and the central limit theorem
- Recognize and apply normal distributions in statistical models

Measurement error

Given a measured quantity X of an unknown true value of a quantity μ, the measurement error is

$$E = X - \mu.$$

If we have repeated measurements (iid) $X_1, X_2, \ldots$, we can compute their mean as our estimator of μ:

$$\bar{X}_n = \frac{1}{n} \sum_i X_i,$$

and the errors are

$$E_i = X_i - \mu.$$

How good is $\bar{X}_n$ as an estimator?

Law of large numbers

Theorem

Let $X_1, X_2, \dots$ be iid random variables with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2 < \infty$. Then, for every $\epsilon > 0$,

$$\lim_{n\to\infty} P\left(|\bar{X}_n - \mu| < \epsilon\right) = 1.$$

i.e. $\bar{X}_n$ **converges in probability** to μ.

A stronger version of the theorem states that

$$P\left(\lim_{n\to\infty} |\bar{X}_n - \mu| < \epsilon\right) = 1.$$

i.e. $\bar{X}_n$ **converges almost surely** to μ.

So, eventually, $\bar{X}_n$ will become a good estimate of μ.

To prove the (weak) law of large numbers, we can use Chebychev's inequality, which states that

$$P\left(|X-\mu| \geq k\sigma\right) \leq \frac{1}{k^2},$$

for any $k > 0$ and any random variable X with mean μ and variance $\sigma^2 < \infty$.

Since $\mathsf{E}(X) = \mu$, $\mathsf{E}(\bar{X}_n) = \mathsf{E}(\frac{1}{n}\sum_i X_i) = \mu$.

Since $\mathsf{Var}(X) = \sigma^2$, $\mathsf{Var}(\bar{X}_n) = \mathsf{Var}(\frac{1}{n}\sum_i X_i) = \frac{\sigma^2}{n}$.

Set $k = \frac{\epsilon\sqrt{n}}{\sigma}$. By Chebychev's inequality, for every $\epsilon > 0$,

$$P\left(|\bar{X}_n - \mu| \geq k\sqrt{\frac{\sigma^2}{n}}\right) \leq \frac{1}{k^2}$$

$$P\left(|\bar{X}_n - \mu| \geq \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2}.$$

Therefore,

$$\lim_{n\to\infty} P\left(|\bar{X}_n - \mu| \geq \epsilon\right) \leq \lim_{n\to\infty} \frac{\sigma^2}{n\epsilon^2} = 0.$$

Finally,

$$\lim_{n\to\infty} P\left(|\bar{X}_n - \mu| < \epsilon\right) = \lim_{n\to\infty} 1 - P\left(|\bar{X}_n - \mu| \geq \epsilon\right) = 1. \quad \blacksquare$$

Assumptions about error

From real-world measurements, scientists have noticed that measurement errors tends to follow two assumptions:

1. Errors are symmetrically distributed about zero.
2. Small errors occur more frequently than large errors.

Carl Gauss used these assumptions and calculus to discover the normal distribution.

Normal distribution

Definition

A random variable X has a normal (Gaussian) distribution if

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

where μ is the mean parameter and τ is the inverse-variance parameter (also known as the precision).

Deriving the normal distribution

Given repeated iid measurements X_i, which have measurement errors

$$E_i = X_i - \mu.$$

These errors are iid:

$$E_i \sim f,$$

where f is unknown, and we **want** to derive it.

To do so, we will determine its mathematical properties based on some assumptions.

Assuming symmetric errors, f is even:

$$f(-e) = f(e), \quad f'(-e) = -f'(e).$$

We will assume that we can estimate the true value μ using the maximum likelihood method:

$$\hat{\mu} = \underset{m}{\operatorname{argmax}}\, f_{\boldsymbol{E}}(\boldsymbol{e} \mid m)$$

Since errors E_i are iid,

$$f_{\boldsymbol{E}}(\boldsymbol{e}) = \prod_i f(e_i) = \prod_i f(x_i - \mu)$$

where we abbreviate the conditioning on m.

We know from the law of large numbers, $\bar{X}$ will be a good estimate for μ as $n \to \infty$. Therefore, we assume that

$$\underset{m}{\operatorname{argmax}}\, f_{\boldsymbol{E}}(\boldsymbol{e} \mid m) = \bar{x}.$$

Since $\log(\cdot)$ is a montone function,

$$\underset{m}{\operatorname{argmax}} \prod_i f(e_i) = \underset{m}{\operatorname{argmax}} \sum_i \log f(e_i).$$

So, we will maximize the log likelihood

$$\ell(m) = \sum_i \log f(e_i).$$

by taking derivative and finding the root

$$\frac{d\,\ell(m)}{de} = \sum_i \frac{f'(e_i)}{f(e_i)} = 0.$$

Define

$$g(e) = \frac{f'(e)}{f(e)}.$$

So,

$$\begin{aligned}
0 &= \sum_i g(e_i) \\
&= \sum_i g(x_i - m) \quad \text{(definition of } e_i\text{)} \\
&= \sum_i g(x_i - \bar{x}) \quad (\bar{x} \text{ is maximum likelihood estimator}) \quad (1)
\end{aligned}$$

Equation (1) holds for any measurements X_i.

Suppose, for some constant b,

$$x_1 = \mu, \quad x_i = \mu - nb,\ i > 1$$

Then,

$$\bar{x} = \mu - (n-1)b$$

Substituting this back into (1),

$$\begin{aligned}
0 &= \sum_i g[x_i - (\mu - (n-1)b)] \\
&= g[\mu - (\mu - (n-1)b)] + \sum_{i>1} g[(\mu - nb) - (\mu - (n-1)b)] \\
&= g[(n-1)b] + \sum_{i>1} g[-b] \\
&= g[(n-1)b] + (n-1)g[-b]
\end{aligned}$$

Therefore,

$$g[(n-1)b] = -(n-1)g[-b]$$

where

$$\begin{aligned} g[-b] &= \frac{f'[-b]}{f[-b]} = \frac{-f'[b]}{f[b]} \quad (f \text{ is even}) \\ &= -g[b] \end{aligned}$$

So, now

$$g[(n-1)b] = (n-1)g[b].$$

This means that g has a linear form:

$$g(e) = k\,e$$

Integrating both sides,

$$\int g(e) = \int k\, e\, de$$
$$\int \frac{f'(e)}{f(e)} = \int k\, e\, de$$
$$\log f(e) = \frac{k}{2} e^2 + c$$

Exponentiating both sides,

$$f(e) = \exp\left(\frac{k}{2} e^2 + c\right) = C \exp\left(\frac{k}{2} e^2\right)$$

Since we assume that smaller errors are more likely, $k < 0$. So, let us define

$$\tau = -k.$$

Now, we need to solve for constant $C > 0$.

Since f is a probability density function, it must satisfy unit measure:

$$\int_{-\infty}^{\infty} f(e)\, de = 1$$

$$\int_{-\infty}^{\infty} \exp\left(-\frac{\tau}{2} e^2\right)\, de = \frac{1}{C}$$

We know that, for $a > 0$,

$$\int_{-\infty}^{\infty} \exp\left(-ax^2\right)\, dx = \sqrt{\frac{\pi}{a}}.$$

Therefore,

$$C = \sqrt{\frac{\tau}{2\pi}}$$

Finally, we have derived the pdf of E_i as

$$f_E(e) = \sqrt{\frac{\tau}{2\pi}} \exp\left(-\frac{\tau}{2}e^2\right),$$

which is a normal distribution centered at 0.

Since $E_i = X_i - \mu$, we can perform a variable transformation to get the distribution of X_i

$$f_X(x) = \sqrt{\frac{\tau}{2\pi}} \exp\left(-\frac{\tau}{2}(x-\mu)^2\right). \quad \blacksquare$$

Remark: Variable transformation also involves the Jacobian (which happens to be 1 here), and we will be covered this later.

Central limit theorem

Theorem

Let $X_1, X_2, \ldots$ be a sequence of iid random variables with $\text{E}(X_i) = \mu$ and $0 < \text{Var}(X_i) = \sigma^2 < \infty$. Define

$$Z = \frac{(\bar{X}_n - \mu)}{\sigma/\sqrt{n}}.$$

Then, for any z,

$$\lim_{n \to \infty} F_Z(z) = \Psi(z),$$

where $\Psi(z)$ is the cdf of Normal$(0, 1)$ given by

$$\Psi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) \, dt.$$

In other words, $F_Z(z)$ **converges in distribution** to Normal$(0, 1)$.

Summary

"It is not knowledge, but the act of learning, not the possession of but the act of getting there, which grants the greatest enjoyment."
- Carl Friedrich Gauss

Casella & Berger 2002, sections 5.5

https://notarocketscientist.xyz/posts/2023-01-27-how-gauss-derived-the-normal-distribution/

Intended learning outcomes

- Derive the normal distribution
- Explain and apply the Chebychev's inequality, laws of large numbers, and the central limit theorem
- Recognize and apply normal distributions in statistical models