Normal Distribution — Modules

25. Definition and Fun Facts

26. Normal Probabilities

27. Central Limit Theorem

28. Bivariate Normal Distribution

5.25 Normal Distribution Definition and Fun Facts

Definition

Fun Facts

Additive Property

Corollary and Standardization

Definition: X has the **normal distribution** with parameters μ and σ^2 if it has p.d.f.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right], \quad \forall x \in \Re.$$

Notation: $X \sim \text{Nor}(\mu, \sigma^2)$

$f(x)$ is "bell-shaped" and symmetric around $x = \mu$, with tails falling off quickly as you move away from μ.

Small σ^2 corresponds to a "tall, skinny" bell curve; large σ^2 gives a "short, fat" bell curve.

Remark: The Normal distribution is also called the Gaussian distrn.

Examples: Heights, weights, SAT scores, crop yields, and averages of things tend to be normal.

Fun Fact (1): $\int_{\Re} f(x)\,dx = 1$.

Proof: Transform to polar coordinates. Good luck.

Fun Fact (2): The c.d.f. is

$$F(x) \;=\; \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(t-\mu)^2}{2\sigma^2}\right] dt \;= ??$$

Remark: No closed-form solution for this. Stay tuned.

Fun Fact (3): $\mathsf{E}[X] = \mu$.

Proof: Integration by parts or m.g.f. (below).

Fun Fact (4): $\mathsf{Var}(X) = \sigma^2$.

Proof: Integration by parts or m.g.f. (below).

Fun Fact (5): The m.g.f. is $M_X(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$.

Proof: Calculus (or look it up in a table of integrals).

Theorem (Additive property of normals): If $X_1, \ldots, X_n$ are *indep* with $X_i \sim \text{Nor}(\mu_i, \sigma_i^2)$, $i = 1, \ldots, n$, then

$$Y \equiv \sum_{i=1}^{n} a_i X_i + b \sim \text{Nor}\Big(\sum_{i=1}^{n} a_i \mu_i + b, \sum_{i=1}^{n} a_i^2 \sigma_i^2\Big).$$

So a linear combination of indep normals is itself normal.

Proof: Since Y is a linear function,

$$
\begin{aligned}
M_Y(t) &= M_{\sum_i a_i X_i + b}(t) = e^{tb} M_{\sum_i a_i X_i}(t) \\
&= e^{tb} \prod_{i=1}^{n} M_{a_i X_i}(t) \quad (X_i\text{'s indep}) \\
&= e^{tb} \prod_{i=1}^{n} M_{X_i}(a_i t) \quad (\text{m.g.f. of linear fn}) \\
&= e^{tb} \prod_{i=1}^{n} \exp\Big[\mu_i (a_i t) + \frac{1}{2}\sigma_i^2 (a_i t)^2\Big] \quad (\text{normal m.g.f.}) \\
&= \exp\Big[\big(\sum_{i=1}^{n} \mu_i a_i + b\big) t + \frac{1}{2}\big(\sum_{i=1}^{n} a_i^2 \sigma_i^2\big) t^2\Big]. \quad (\text{Done.})
\end{aligned}
$$

Remark: A normal distrn is *completely characterized* by its mean and variance.

By the above, we know that a linear combination of indep normals is still normal.

Therefore, when we add up indep normals, all we have to do is figure out the mean and variance — the normality of the sum comes for free.

Example: $X \sim$ Nor(3, 4), $Y \sim$ Nor(4, 6) and X, Y indep. Find the distrn of $2X - 3Y$.

Solution: This is *normal* with

$$\text{E}[2X - 3Y] = 2\text{E}[X] - 3\text{E}[Y] = 2(3) - 3(4) = -6$$

and

$$\text{Var}(2X - 3Y) = 4\text{Var}(X) + 9\text{Var}(Y) = 70.$$

Thus, $2X - 3Y \sim \text{Nor}(-6, 70)$.

Corollary (of Theorem):

$$X \sim \text{Nor}(\mu, \sigma^2) \Rightarrow aX + b \sim \text{Nor}(a\mu + b, a^2\sigma^2).$$

Proof: Immediate from Theorem after noting that $\text{E}[aX + b] = a\mu + b$ and $\text{Var}(aX + b) = a^2\sigma^2$.

Corollary (of Corollary):

$$X \sim \text{Nor}(\mu, \sigma^2) \Rightarrow Z \equiv \frac{X - \mu}{\sigma} \sim \text{Nor}(0, 1).$$

Proof: Use above Cor with $a = 1/\sigma$ and $b = -\mu/\sigma$.

Definition: The Nor$(0,1)$ distrn is called the **standard normal** distribution.

The standard normal is nice because there are tables available for its c.d.f.

You can standardize any normal RV X into a standard normal by applying the transformation $Z = (X - \mu)/\sigma$. Then you can use the c.d.f. tables.