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.

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

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

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

So a linear combination of indep normals is itself normal.

Proof: Since Y is a linear function,

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

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, $2X - 3Y \sim$ 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.

5.26 Normal Probabilities

Standard Normal Distribution

Examples

Sample Mean of Normal Observations

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

Notation: The Nor$(0,1)$ is often denoted by Z.

The p.d.f. of the Nor$(0,1)$ is

$$\phi(z) \equiv \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad z \in \Re.$$

The c.d.f. is

$$\Phi(z) \equiv \int_{-\infty}^{z} \phi(t)\, dt, \quad z \in \Re.$$

Remarks:

$$\Pr(Z \le a) \;=\; \Phi(a)$$
$$\Pr(Z \ge b) \;=\; 1 - \Phi(b)$$
$$\Pr(a \le Z \le b) \;=\; \Phi(b) - \Phi(a)$$

$$\Phi(0) \;=\; 1/2$$
$$\Phi(-b) \;=\; \Pr(Z \le -b) \;=\; \Pr(Z \ge b) \;=\; 1 - \Phi(b)$$
$$\Pr(-b \le Z \le b) \;=\; \Phi(b) - \Phi(-b) \;=\; 2\Phi(b) - 1$$

Famous Nor$(0, 1)$ table values.

z	$\Phi(z) = \text{Pr}(Z \leq z)$
0.00	0.5000
1.00	0.8413
1.28	$0.8997 \approx 0.90$
1.645	0.9500
1.96	0.9750
2.33	$0.9901 \approx 0.99$
3.00	0.9987
4.00	≈ 1.0000

Famous **Inverse** $\text{Nor}(0,1)$ table values.

$\Phi^{-1}(p)$ is the value of z such that $\Phi(z) = p$.

p	$\Phi^{-1}(p)$
0.90	1.28
0.95	1.645
0.975	1.96
0.99	2.33
0.995	2.58

Example: $X \sim \text{Nor}(21, 4)$. Find $\Pr(19 < X < 22.5)$.

Standardizing, we get

$$
\begin{aligned}
&\Pr(19 < X < 22.5) \\
&= \Pr\left(\frac{19-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{22.5-\mu}{\sigma}\right) \\
&= \Pr\left(\frac{19-21}{2} < Z < \frac{22.5-21}{2}\right) \\
&= \Pr(-1 < Z < 0.75) \\
&= \Phi(0.75) - \Phi(-1) \\
&= \Phi(0.75) - [1 - \Phi(1)] \\
&= 0.7734 - [1 - 0.8413] = 0.6147.
\end{aligned}
$$

Example: Suppose that

Heights of men are $M \sim$ Nor$(68, 4)$ and

Heights of women are $W \sim$ Nor$(65, 1)$.

Select a man and woman *independently* at random.

Find the probability that the woman is taller than the man.

Note that

$$\begin{aligned} W - M &\sim \text{Nor}(\text{E}[W-M], \text{Var}(W-M)) \\ &\sim \text{Nor}(65-68, 1+4) \sim \text{Nor}(-3, 5). \end{aligned}$$

Then

$$\begin{aligned} \Pr(W > M) &= \Pr(W - M > 0) \\ &= \Pr\left(Z > \frac{0+3}{\sqrt{5}}\right) \\ &= 1 - \Phi(3/\sqrt{5}) \\ &\approx 1 - 0.91 = 0.09. \end{aligned}$$

Sample Mean of Normal Observations

The sample mean of $X_1, \ldots, X_n$ is $\bar{X} \equiv \sum_{i=1}^{n} X_i/n$.

Corollary (of old Theorem): $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} \text{Nor}(\mu, \sigma^2) \Rightarrow \bar{X} \sim \text{Nor}(\mu, \sigma^2/n)$.

Proof: By previous work, as long as $X_1, \ldots, X_n$ are i.i.d. something, we have $\text{E}[\bar{X}] = \mu$ and $\text{Var}(\bar{X}) = \sigma^2/n$. Since $\bar{X}$ is a linear combination of independent normals, it's also **normal**. Done.

Remark: This result is *very significant!* As the number of observations increases, Var($\bar{X}$) gets *smaller* (while E[$\bar{X}$] remains constant).

In the upcoming statistics portion of the course, we'll learn that this makes the sample mean $\bar{X}$ an excellent **estimator** for the mean μ, which is typically unknown in practice.

Example: Suppose that $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} \text{Nor}(\mu, 16)$. Find the sample size n such that

$$\Pr(|\bar{X} - \mu| \leq 1) \;\geq\; 0.95.$$

How many observations should you take so that $\bar{X}$ will have a good chance of being close to μ?

Solution: Note that $\bar{X} \sim \text{Nor}(\mu, 16/n)$. Then...

$$\begin{aligned}
\Pr(|\bar{X} - \mu| \le 1) &= \Pr(-1 \le \bar{X} - \mu \le 1) \\
&= \Pr\left(\frac{-1}{4/\sqrt{n}} \le \frac{\bar{X} - \mu}{4/\sqrt{n}} \le \frac{1}{4/\sqrt{n}}\right) \\
&= \Pr\left(\frac{-\sqrt{n}}{4} \le Z \le \frac{\sqrt{n}}{4}\right) \\
&= 2\Phi(\sqrt{n}/4) - 1.
\end{aligned}$$

Now we have to find n such that this probability is at least 0.95. . . .

$2\Phi(\sqrt{n}/4) - 1 \geq 0.95$ iff

$\Phi(\sqrt{n}/4) \geq 0.975$ iff

$$\frac{\sqrt{n}}{4} \geq \Phi^{-1}(0.975) = 1.96$$

iff $n \geq 61.47$ or 62.

So if you take the average of 62 observations, then $\bar{X}$ has a 95% chance of being within 1 of μ.

5.27 Central Limit Theorem

CLT

Example

Normal Approximation to the Binomial

The most important theorem in prob and stats.

Central Limit Theorem: Suppose $X_1, \ldots, X_n$ are i.i.d. with $\mathrm{E}[X_i] = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$. Then as $n \to \infty$,

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{\mathcal{D}} \mathrm{Nor}(0,1),$$

where "$\xrightarrow{\mathcal{D}}$" means that the c.d.f. $\to$ the Nor$(0,1)$ c.d.f.

Proof: Not in this class.

Remarks: (1) So if n is large, then $\bar{X} \approx \text{Nor}(\mu, \sigma^2/n)$.

(2) The X_i's *don't have to be normal* for the CLT to work!

(3) You usually need $n \geq 30$ observations for the approximation to work well. (Need fewer observations if the X_i's come from a symmetric distribution.)

(4) You can almost always use the CLT if the observations are i.i.d.

Example: Suppose $X_1, \ldots, X_{100} \overset{\text{iid}}{\sim} \text{Exp}(1/1000)$. Find $\Pr(950 \leq \bar{X} \leq 1050)$.

Solution: Recall that if $X_i \sim \text{Exp}(\lambda)$, then $\text{E}[X_i] = 1/\lambda$ and $\text{Var}(X_i) = 1/\lambda^2$.

Further, if $\bar{X}$ is the sample mean based on n observations, then

$$\text{E}[\bar{X}] \;=\; \text{E}[X_i] \;=\; 1/\lambda \quad \text{and}$$

$$\text{Var}(\bar{X}) \;=\; \text{Var}(X_i)/n \;=\; 1/(n\lambda^2).$$

For our problem, $\lambda = 1/1000$ and $n = 100$, so that $\mathsf{E}[\bar{X}] = 1000$ and $\mathsf{Var}(\bar{X}) = 10000$.

So by the CLT,

$$\begin{aligned} &\Pr(950 \le \bar{X} \le 1050) \\ &= \Pr\left(\frac{950 - \mathsf{E}[\bar{X}]}{\sqrt{\mathsf{Var}(\bar{X})}} \le \frac{\bar{X} - \mathsf{E}[\bar{X}]}{\sqrt{\mathsf{Var}(\bar{X})}} \le \frac{1050 - \mathsf{E}[\bar{X}]}{\sqrt{\mathsf{Var}(\bar{X})}}\right) \\ &\approx \Pr\left(\frac{950 - 1000}{100} \le Z \le \frac{1050 - 1000}{100}\right) \\ &\approx \Pr\left(-\frac{1}{2} \le Z \le \frac{1}{2}\right) = 2\Phi(1/2) - 1 = 0.383. \end{aligned}$$

Example: Suppose $X_1, \ldots, X_{100}$ are i.i.d. from some distribution with mean 1000 and standard deviation 1000. Find $\Pr(950 \le \bar{X} \le 1050)$.

Solution: By exactly the same manipulations as in the previous example, the answer ≈ 0.383.

Notice that we didn't care whether or not the data came from an exponential distrn. We just needed the mean and variance.

Normal Approximation to the Binomial(n, p)

Suppose $Y \sim \text{Bin}(n, p)$, where n is very large. In such cases, we usually approximate the Binomial via an appropriate Normal distribution.

The CLT applies since $Y = \sum_{i=1}^{n} X_i$, where the X_i's are i.i.d. Bern(p).

Then

$$\frac{Y - \text{E}[Y]}{\sqrt{\text{Var}(Y)}} = \frac{Y - np}{\sqrt{npq}} \approx \text{Nor}(0, 1).$$

Why do we need such an approximation?

Example: Suppose $Y \sim \text{Bin}(100, 0.8)$ and we want

$$\Pr(Y \geq 90) = \sum_{i=90}^{100} \binom{100}{i} (0.8)^i (0.2)^{100-i}.$$

Good luck with the binomial coefficients (they're too big) and number of terms to sum up (it's going to get tedious). I'll come back to visit you in an hour.

So how do we use the approximation?

Example: The Braves play 100 indep baseball games, each of which they have prob 0.8 of winning. What's the prob that they win at least 90?

$Y \sim \text{Bin}(100, 0.8)$ and we want $\Pr(Y \geq 90)$ (as in the last example)...

$$\Pr(Y \geq 90) \;=\; \Pr(Y \geq 89.5) \quad (\text{“continuity correction”})$$

$$\approx\; \Pr\left(Z \geq \frac{89.5 - np}{\sqrt{npq}}\right) \quad (\text{CLT})$$

$$=\; \Pr\left(Z \geq \frac{89.5 - 80}{\sqrt{16}}\right) \;=\; \Pr(Z \geq 2.375) \;=\; 0.0088.$$

Use the continuity correction since the Binomial is a discrete distrn while the Normal is cts. If you don't want to use it, don't worry too much.

5.28 Bivariate Normal and Friends

Bivariate Normal Distrn

Bivariate Normal Example

Lognormal Distrn

Simulating Normal RV's

(X, Y) has the **Bivariate Normal Distrn** if it has p.d.f.

$$f(x, y) = C \exp\left\{\frac{-\left[z_X^2(x) - 2\rho z_X(x) z_Y(y) + z_Y^2(y)\right]}{2(1-\rho^2)}\right\}$$

where

$$\rho \equiv \text{Corr}(X, Y), \quad C \equiv \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}},$$

$$z_X(x) \equiv \frac{x - \mu_X}{\sigma_X} \text{ and } z_Y(y) \equiv \frac{y - \mu_Y}{\sigma_Y}.$$

Pretty nasty joint p.d.f., eh?

In fact, $X \sim \text{Nor}(\mu_X, \sigma_X^2)$ and $Y \sim \text{Nor}(\mu_Y, \sigma_Y^2)$.

Example: (X, Y) could be a person's (height,weight). The two quantities are marginally normal, but positively correlated.

If you want to calculate bivariate normal probabilities, you'll need to evaluate quantities like

$$\mathsf{Pr}(a < X < b, c < Y < d) \;=\; \int_c^d \int_a^b f(x,y)\,dx\,dy,$$

which will probably require numerical integration techniques.

Fun Fact (which will come up later when we discuss regression): The conditional distribution of Y given that $X = x$ is also normal. In particular,

$$Y|X = x \sim \text{Nor}(\mu_Y + \rho(\sigma_Y/\sigma_X)(x - \mu_X), \sigma_Y^2(1 - \rho^2)).$$

Information about X helps to update the distribution of Y.

Example: Consider students at a university. Let X be their combined SAT scores (Math and Verbal), and Y their freshman GPA (out of 4). Suppose a study reveals that

$$\mu_X = 1300, \quad \mu_Y = 2.3,$$

$$\sigma_X^2 = 6400, \quad \sigma_Y^2 = 0.25, \quad \rho = 0.6.$$

Find $\Pr(Y \geq 2 | X = 900)$.

First,

$$\begin{aligned} &\mathsf{E}[Y|X = 900] \\ &= \mu_Y + \rho(\sigma_Y/\sigma_X)(x - \mu_X) \\ &= 2.3 + \rho(\sqrt{0.25/6400})(900 - 1300) = 0.8, \end{aligned}$$

indicating that the expected GPA of a kid with 900 SAT's will be 0.8.

Second,

$$\mathsf{Var}(Y|X = 900) = \sigma_Y^2(1 - \rho^2) = 0.16.$$

Thus,

$$Y|X = 900 \;\sim\; \text{Nor}(0.8, 0.16).$$

Now we can calculate

$$\begin{aligned}\Pr(Y \geq 2|X = 900) \;&=\; \Pr\left(Z \geq \frac{2 - 0.8}{\sqrt{0.16}}\right) \\ &=\; 1 - \Phi(3) \;=\; 0.0013.\end{aligned}$$

This guy doesn't have much chance of having a good GPA.

Lognormal Distrn

Definition: If $Y \sim \text{Nor}(\mu_Y, \sigma_Y^2)$, then $X \equiv e^Y$ has the **lognormal distrn** with parameters (μ_Y, σ_Y^2).

Turns Out: The p.d.f. of the lognormal is

$$f(x) = \frac{1}{x\sigma_Y\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma_Y^2}[\ell\text{n}(x) - \mu_Y]^2\right\}, \quad x > 0.$$

Further,

$$\mathsf{E}[X] = \exp\left\{\mu_Y + \frac{\sigma_Y^2}{2}\right\}$$

$$\mathsf{Var}(X) = \exp\{2\mu_Y + \sigma_Y^2\}\Big(\exp\{\sigma_Y^2\} - 1\Big)$$

The lognormal distrn has lots of other nice properties.

See any good prob/stats book for details.

Example: Suppose $Y \sim \text{Nor}(10, 4)$ and let $X = e^Y$. Then

$$\begin{aligned} \Pr(Y \le \ell\text{n}(1000)) &= \Pr\left(Z \le \frac{\ell\text{n}(1000) - 10}{2}\right) \\ &= \Phi(-1.55) = 0.061. \end{aligned}$$

Simulating Normal RV's

How would you generate a normal RV's when conducting computer simulation experiments?

Theorem (Box and Müller): If $U_1, U_2 \stackrel{\text{iid}}{\sim} \text{U}(0,1)$, then

$$\begin{aligned} Z_1 &= \sqrt{-2\ell\text{n}(U_1)}\,\cos(2\pi U_2), \\ Z_2 &= \sqrt{-2\ell\text{n}(U_1)}\,\sin(2\pi U_2) \end{aligned}$$

are iid Nor(0,1).

Remarks: (1) Proof: Not here.

(2) Many other ways to generate Nor(0,1)'s, but this is the easiest.

(3) Cosine and Sine must be calculated in *radians*, not degrees.

(4) To get $X \sim$ Nor(μ, σ^2), just take $X = \mu + \sigma Z$.