Normal Distribution

Standard Normal Distribution, $N(0, 1)$

p.d.f.:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

m.g.f.:

$$\phi(t) = \mathbb{E}(e^{tX}) = e^{t^2/2}$$

Proof - Simplify integral by completing the square:

$$\phi(t) = \int e^{tx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int e^{tx-x^2/2} dx =$$

$$\frac{1}{\sqrt{2\pi}} \int e^{t^2/2 - t^2/2 + tx - x^2/2} dx = \frac{1}{\sqrt{2\pi}} e^{t^2/2} \int e^{-\frac{1}{2}(t-x)^2} dx$$

Then, perform the change of variables $y = x - t$:

$$= \frac{1}{\sqrt{2\pi}} e^{t^2/2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} dy = e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} dy = e^{t^2/2} \int f(x) dx = e^{t^2/2}$$

Use the m.g.f. to find expectation of $X$ and $X^2$ and therefore Var($X$):

$$\mathbb{E}(X) = \phi'(0) = te^{t^2/2}|_{t=0} = 0; \mathbb{E}(X^2) = \phi''(0) = e^{t^2/2}t^2 + e^{t^2/2}|_{t=0} = 1; \text{Var}(X) = 1$$

Consider $X \sim N(0, 1), Y = \sigma X + \mu$, find the distribution of $Y$:

$$\mathbb{P}(Y \leq y) = \mathbb{P}(\sigma X + \mu \leq y) = \mathbb{P}(X \leq \frac{y-\mu}{\sigma}) = \int_{-\infty}^{\frac{y-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

p.d.f. of $Y$:

$$f(y) = \frac{\partial \mathbb{P}(Y \leq y)}{\partial y} = \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{y-\mu}{\sigma}\right)^2} \frac{1}{\sigma} = \frac{1}{\sigma \sqrt{2\pi}} e^{-\left(\frac{y-\mu}{\sigma}\right)^2} \rightarrow N(\mu, \sigma)$$

$$\mathbb{E}Y = \mathbb{E}(\sigma X + \mu) = \sigma(0) + \mu(1) = \mu$$

$$\mathbb{E}(Y - \mu)^2 = \mathbb{E}(\sigma X + \mu - \mu)^2 = \sigma^2 \mathbb{E}(X^2) = \sigma^2 - \text{variance of } N(\mu, \sigma)$$

$$\sigma = \sqrt{\text{Var}(X)} \rightarrow \text{standard deviation}$$
To describe an altered standard normal distribution $N(0, 1)$ to a normal distribution $N(\mu, \sigma)$, the peak is located at the new mean $\mu$, and the point of inflection occurs $\sigma$ away from $\mu$.

Moment Generating Function of $N(\mu, \sigma)$:

$Y = \sigma X + \mu$

$$\phi(t) = E^tY = E^t(\sigma X + \mu) = E^t(\sigma X) e^{t\mu} = e^{t\mu}E^tX = e^{t\mu}e^{t(\sigma)^2/2} = e^{t\mu + t^2(\sigma)^2/2}$$

Note: $X_1 \sim N(\mu_1, \sigma_1), ..., X_n \sim N(\mu_n, \sigma_n)$ - independent.

$Y = X_1 + ... + X_n$, distribution of $Y$:

Use moment generating function:

$$E^tY = E^t(X_1 + ... + X_n) = E^tX_1 ... E^tX_n = e^{t\mu_1 + \sigma_1^2t^2/2} ... e^{t\mu_n + \sigma_n^2t^2/2}$$

The sum of different normal distributions is still normal!
This is not always true for other distributions (such as exponential).

Example:

$X \sim N(\mu, \sigma), Y = cX$, find that the distribution is still normal:

$Y = e(c(\sigma N(0, 1) + \mu) = (c\sigma)N(0, 1) + (c\mu)$

$Y \sim eN(\mu, \sigma) = N(c\mu, c\sigma)$

Example:

$Y \sim N(\mu, \sigma)$

$\mathbb{P}(a \leq Y \leq b) = \mathbb{P}(a \leq \sigma x + \mu \leq b) = \mathbb{P} \left( \frac{a - \mu}{\sigma} \leq X \leq \frac{b - \mu}{\sigma} \right)$

This indicates the new limits for the standard normal.

Example:

Suppose that the heights of women: $X \sim N(65, 1)$ and men: $Y \sim N(68, 2)$

$\mathbb{P}(X > Y) = \mathbb{P}(X - Y > 0)$

$Z = X - Y \sim N(65 - 68, \sqrt{1^2 + 2^2}) = N(-3, \sqrt{5})$

$\mathbb{P}(Z > 0) = \mathbb{P} \left( \frac{Z - (-3)}{\sqrt{5}} > \frac{z - (-3)}{\sqrt{5}} \right) = \mathbb{P} \left( \text{standard normal} > \frac{3}{\sqrt{5}} \right) = 0.09$

Probability values tabulated in the back of the textbook.

**Central Limit Theorem**

Flip 100 coins, expect 50 tails, somewhere 45-50 is considered typical.
Flip 10,000 coins, expect 5,000 tails, and the deviation can be larger, perhaps 4,950-5,050 is typical.

\[ X_i = \{1(\text{tail}); 0(\text{head})\} \]

\[
\frac{\text{number of tails}}{n} = \frac{X_1 + \ldots + X_n}{n} \quad \rightarrow \quad \mathbb{E}(X_1) = \frac{1}{2} \quad \text{by LLN} \quad \text{Var}(X_1) = \frac{1}{2}(1 - \frac{1}{2}) = \frac{1}{4}
\]

But, how do you describe the deviations? 
\( X_1, X_2, \ldots, X_n \) are independent with some distribution \( P \)
\[
\mu = \mathbb{E}X_1, \sigma^2 = \text{Var}(X_1); \bar{x} = \frac{1}{n} \sum_{i=1}^{n} X_i \rightarrow \mathbb{E}X_1 = \mu
\]

\( \bar{x} - \mu \) on the order of \( \sqrt{n} \rightarrow \frac{\sqrt{n}(\bar{x} - \mu)}{\sigma} \) behaves like standard normal.
\[
\frac{\sqrt{n}(\bar{x} - \mu)}{\sigma} \quad \text{is approximately standard normal} \quad N(0, 1) \quad \text{for large} \ n
\]

\[ \mathbb{P}(\frac{\sqrt{n}(\bar{x} - \mu)}{\sigma} \leq x) \xrightarrow{n \rightarrow \infty} \mathbb{P}(\text{standard normal} \leq x) = N(0, 1)(-\infty, x) \]

This is useful in terms of statistics to describe outcomes as likely or unlikely in an experiment.

\[ \mathbb{P}(\text{number of tails} \leq 4900) = \mathbb{P}(X_1 + \ldots + X_{10,000} \leq 4,900) = \mathbb{P}(\bar{x} \leq 0.49) = \]
\[ = \mathbb{P}\left(\frac{\sqrt{10,000}(\bar{x} - \frac{1}{2})}{\frac{1}{2}} \leq \frac{\sqrt{10,000}(0.49 - 0.5)}{\frac{1}{2}}\right) \approx N(0, 1)(-\infty, -100(0.01) = -2) = 0.0267 \]

Tabulated values always give for positive \( X \), area to the left.

In the table, look up -2 by finding the value for 2 and taking the complement.

** End of Lecture 21 **