Review.

1. Gaussian distribution

- $X$ has Gaussian (normal) distribution, $X \sim N(\mu, \sigma^2)$ if
  \[ f_X(x) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R} \]

  \[ E[X] = \mu, \quad \text{Var}(X) = \sigma^2 \]

- Standardized version of $X$, $\tilde{X} = \frac{X-\mu}{\sigma}$ is distributed as $N(0,1)$.

  \[ F_X(a) = P(X \leq a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) \, du = \frac{1}{2} \left[ 1 + \text{erf}\left(\frac{a-\mu}{\sigma\sqrt{2}}\right) \right] \]

  \[ F_X^c(a) = P(X > a) = \int_{a}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) \, du = 1 - F_X(a) = \Phi(a) \]

- Notice that $\Phi(a) = \Phi(-a)$ because of symmetry.

2. Gaussian approximation with error correction

- Let $S \sim \text{Bin}(np)$ denote a random variable with Bernoulli distribution. Let $X \sim N(\mu, \sigma^2)$ denote a Gaussian random variable with same mean and variance as $S$, i.e., $\mu = np$, $\sigma^2 = np(1-p)$

  \[ P(S < k) \approx P(X < k + 0.5) \]
  \[ P(S > k) \approx P(X > k + 0.5) \]
  \[ P(S \geq k) \approx P(X \geq k - 0.5) \]

3. ML parameter estimation for continuous-type random variables

   (i) pdf of random variable $X$ belongs to a family of parametrized distributions: $f_\theta$

   (ii) We are given an observation $X = u$

   (iii) We want to find the parameter that maximizes the pdf at $u$, i.e., $\hat{\theta}_M = \arg\max_{\theta} f_\theta(u)$

Today:

1. The distribution of a function of a random variable

   - Suppose that $X$ is a continuous-type random variable and $Y = g(X)$ for some function $g$.
   - Our goal is to find distribution of $Y$.

   **Step 1:** Find whether $Y$ is continuous-type or discrete-type and find its support.
Step 1: Find whether $Y$ is continuous-type or discrete-type and find its support.

To do this, use the following steps:

(i) sketch pdf of $X$ and function $g$ on the same plot.
(ii) Identify support of $X$, i.e., region over $x$-axis for which $f_X$ is positive. Let's call this region $A$.
(iii) Identify $g(A)$ on the $y$-axis. This is support of $Y$. If $g(A)$ is itself a continuous region, $Y$ is continuous-type.

If $g(A)$ is union of points, it is discrete type.

Note: Sometimes, Step 1 is easy. You can identify support of $Y$ without using above steps.

Example: Suppose that $Y = X^2$ and $f_X(u) = \frac{1}{2}e^{-\frac{1}{2}u}$ for $u \in \mathbb{R}$. We want to find $f_Y(u)$.

Notice that $Y = g(X)$ where $g(u) = u^2$.

If $Y$ is continuous-type:

Step 2: Find CDF of $Y$ using $F_Y(c) = P(Y \leq c) = P(g(X) \leq c) = \int_{g^{-1}(c)} f_X(u) du$.

Note: If it is possible, write down $F_Y$ in terms of $F_X$. This way you do not need to do integration.

Example: Step 2. So far, we have that support of $Y$ is positive value. For any $c > 0$,

$$F_Y(c) = P(X^2 \leq c) = P(\{X \leq \sqrt{c}\}) = P(X \leq \sqrt{c}) - P(X \geq \sqrt{c}) = F_X(\sqrt{c}) - F_X(-\sqrt{c})$$

Step 3: Find pdf of $Y$ by differentiating $F_Y$, i.e., $f_Y(u) = \frac{df_Y(u)}{du}$.

Example: $\frac{df_Y(u)}{du} = \frac{d}{du} \left[ F_X(\sqrt{u}) - F_X(-\sqrt{u}) \right] = \frac{1}{2\sqrt{u}} f_X(\sqrt{u}) + \frac{1}{2\sqrt{u}} f_X(-\sqrt{u}) = \frac{1}{2\sqrt{u}} e^{-\frac{1}{2}u}$ for $u > 0$.

If $Y$ is discrete-type:

Step 1: Find pmf of $Y$ using $P(Y = k) = P(g(X) = k) = \int f_X(u) du$.

Some remarks:

After finding distribution of $Y$, you can use it to get statistical quantities related to $Y$ using $f_Y$ or LOTUS.

Example: for the above example,

$$E[Y] = E[g(X)] = E[X^2] = \frac{1}{2} \int_{-\infty}^{\infty} u^2 e^{-\frac{1}{2}u} du = \frac{1}{2} \int_{0}^{\infty} u^2 e^{-\frac{1}{2}u} du = 2! \text{ integration by part + LOTUS}$$

$$E[Y] = \int_{-\infty}^{\infty} f_Y(u) du = \int_{0}^{\infty} u \cdot \frac{1}{2\sqrt{u}} e^{-\frac{1}{2}u} du = \int_{0}^{\infty} \frac{1}{2\sqrt{u}} e^{-\frac{1}{2}u} du = \int_{0}^{\infty} e^v dv = 2! \text{ change of variable } u \rightarrow v$$

Although we assumed $X$ is continuous-type, the same can be used if $X$ is a discrete random variable.
Although we assumed $X$ is continuous-type, the same can be used if $X$ is a discrete random variable.

**Example 3.8.11** Suppose $X$ is a continuous-type random variable with CDF $F_X$. Let $Y$ be the result of applying $F_X$ to $X$, that is, $Y = F_X(X)$. Find the distribution of $Y$.

**Solution:**

**Step 1.** Notice that $Y = g(X) = F_X(X)$. Hence $g = F_X$. We know range of $F_X$ is $[0,1]$ and $X$ is continuous type. Hence, support of $Y$ is $[0,1]$.

**Step 2.** $P(Y < v) = P(F_X(X) < v), \forall v \in [0,1]$.

Since $X$ is continuous type, $F_X$ is increasing. Moreover, for any $v \in (0,1)$ there exists a unique $c_v$ such that $F_X(c_v) = v$. Hence, for any $v \in (0,1)$ we have $P(F_X(X) < v) = P(X < c_v) = F_X(c_v) = v$.

Notice that $P(Y \leq 0) = 0$ and $P(Y \leq 1) = 1$.

**Step 3.** $f_Y(v) = \frac{df_{F_X}(v)}{dv} = 1 \quad \text{for } v \in [0,1]$.

Hence $Y$ is a uniformly distributed random variable over $[0,1]$.

**Remark:**

Suppose that $Y \sim \text{Unif}([0,1])$ & $F$ is cdf of a continuous type random variable. The above example suggests $X = F^{-1}(Y)$ is a continuous type random variable with cdf $F_X$. Hence, if we have a method for generating $Y \sim \text{Unif}([0,1])$ random variable, we can generate any continuous-type random variable with cdf $F$, by pass $Y$ through $F^{-1}$. 