The distribution of a function of a random variable

Suppose that $Y = g(X)$. $X$ is continuous type. We want to find the distribution of $Y$.

**Step 1.** Identify type of $Y$ (continuous type or discrete type) and support of $Y$ (w.r.t. $f_X(u)$).

If $Y$ is continuous type,

**Step 2.** Find its cdf. $F_Y(c) = P(Y \leq c) = P(g(X) \leq c) = \int_{u : g(u) \leq c} f_X(u) \, du$.

**Step 3.** Take derivative of $F_Y(c)$ to derive pdf of $Y$.

$$f_Y(c) = \frac{dF_Y(c)}{dc}$$

If $Y$ is discrete type,

**Step 2.** Calculate pmf of $Y$. $P(Y = k) = P(g(X) = k) = \int_{u : g(u) = k} f_X(u) \, du$.

**Important example.**

$X$ is continuous type with CDF $F_X$. $Y = F_X(X)$ is uniformly distributed between 0 and 1.

**Today.**

- Generating random variable with a specified distribution.
- Sanity hypothesis testing with continuous type observations.

- Generating random variable with a specified distribution.

Recall that a random variable $X$ is a function from $\Omega$ to $\mathbb{R}$, i.e., $X: \Omega \rightarrow \mathbb{R}$. For any $\omega \in \Omega$, $X(\omega)$ is a realization or an instance of underlying experiment:

```
experiment (Ω, F, P) \rightarrow Generates \ w \rightarrow Random \ variable \ X \rightarrow Generates \ X(\omega)
```

Recall that we only care about statistical properties, so mapping outcomes of experiment to $\mathbb{R}$ using random variables makes it possible to focus on what matters, i.e., distributions.

We also studied famous distributions that are common in practice. Now what if we want to run a numerical simulation via computer? How should we generate an instance $X(\omega)$, given CDF of $X$ is $F_X$?

(i). If $X$ is continuous type, then $F_X$ is increasing, and $Y = F_X(X)$ has $\text{Unif}([0,1])$ distribution. Using this, if $U$ is a uniform r.v. over $[0,1]$, $X = F_X^{-1}(U)$ is a random variable with CDF $F_X$. Notice that graphically...
\[ F_X^* \] is resulted by reflection of \( F \) via the line passing through origin with slope 1.

Notice that for any \( x \in (a, b) \) there exist a unique \( c_x \) for which \( F_X(c_x) = x \). Hence, \( F_X^* \) is well defined over \((0,1)\).

Hence, if \( F \) is cdf of a continuous-type random variable, and \( U \) is uniformly distributed over \([0,1]\), then \( X = F(U) \) has cdf \( F_X = F \).

This suggests to generate a random variable with arbitrary distribution, we can generate a uniform random variable \( U \) & pass it through some function \( g \) s.t. \( X = g(U) \) has the desired distribution.

(ii). Consider a function \( F: \mathbb{R} \to [0,1] \) and suppose that \( F \) is a cdf, i.e.,

1. \( F \) is increasing,
2. \( \lim_{c \to -\infty} F(c) = 0, \lim_{c \to \infty} F(c) = 1 \)
3. \( F \) is right continuous.

Our goal is to find a function \( g \) s.t. the cdf of \( X = g(U) \) is given by \( F \), where \( U \sim \text{Unif}(0,1) \).

We observed that if \( F \) is cdf of a continuous-type random variable, then \( g = F^{-1} \) is the function we are interested in. Recall that \( F^{-1} \) is reflection of \( F \). Let us naively do the same for general cdf function.

The green line is not a function, so let's make it into a function with domain \([0,1]\) by

(i) removing vertical line (ii) replacing dashed horizontal line with solid line

\[ F_X^* \] reflection of \( F_X \) via the line passing through origin with slope one.

Notice that values at jumps.
You can check that the function $F^{-1}$ resulted from the above procedure is

\[ F^{-1}(u) = \min \{ c : F_X(c) \geq u \} \quad \text{for all } u \in [0,1] \]

Suppose that $c_u = F^{-1}(u) = \min \{ c : F(c) > u \}$.

We have: $F^{-1}(u) < a$ if and only if $c_u < a$ if and only if $u \leq F(a)$.

Hence, if we define $X = F^{-1}(U)$ then cdf of $X$ is given by:

\[ F_X(c) = P(X \leq c) = P(F^{-1}(U) \leq c) \]

By (9) $F^{-1}(U) \leq c$ if and only if $U \leq F(c)$.

\[ F_X(c) = P(F^{-1}(U) \leq c) = P(U \leq F(c)) = F(c) \]

since for a uniformly distributed random variable $U$ over $[0,1]$, $F_U(a) = P(U \leq a) = a$ if $a \in [0,1]$.

So, we have the following statement:

Suppose that $F$ is a valid CDF. Define $F^{-1}(u) = \min \{ c : F(c) > u \}$ for any $u \in (0,1]$.

Suppose that $U$ is a uniformly distributed random variable over $[0,1]$. Define $X = F^{-1}(U)$.

Then cdf of $X$ is $F$, i.e., $F_X(c) = P(X \leq c) = F(c)$.

**Question. How to find $F^{-1}$?**

**approach 1.** In some cases, $F$ is invertible; for values in $(0,1)$, i.e., for any $u \in (0,1)$ there exists a unique $c_u$ st. $F(c_u) = u$. In these cases, directly calculate $F^{-1}$.

**approach 2.** In some cases graphs are easier to plot. In these cases:

(i) reflect $F$ via the line that passes through origin with slope 1.

(ii) remove vertical lines.

(iii) make the resulted graph to a proper functions with domain $[0,1]$ by replacing dashed horizontal lines with solid lines.
Example.

3.37. [Generation of random variables with specified probability density function]
Find a function \( g \) so that, if \( U \) is uniformly distributed over the interval \([0, 1]\), and \( X = g(U) \), then \( X \) has the pdf:

\[
f_X(u) = \begin{cases} 
2u & \text{if } 0 \leq u \leq 1 \\
0 & \text{else.}
\end{cases}
\]

**Solution:** Notice that \( g' = F_X' \). So, the first step is to find \( F_X \).

\[
F_X(c) = P(X \leq c) = \begin{cases} 
1 & \text{if } c > 1 \\
c^2 & \text{if } c \in [0,1] \\
0 & \text{if } c < 0
\end{cases}
\]

Notice that the inverse of \( F_X \), for \( u \in (0,1) \) is given by \( F_X^{-1}(u) = \sqrt{u} \). Hence, the function \( g \) is given by \( g(u) = \sqrt{u} \) with domain \( \geq [0,1] \).

Note: value of \( g(0) \) and \( g(1) \) does not matter since \( P(U=0) = P(U=1) = 0 \).

Problem. Find a function \( g \) st. if \( U \) is uniformly distributed over \([0,1]\), and \( X = g(U) \), then \( F_X \) is given by:

\[
\begin{array}{c|c|c}
\hline
u & 0.5 & 1 \\
\hline
0 & 0 & 1 \\
0.5 & 1 & 1 \\
1 & 1 & 1 \\
\hline
\end{array}
\]

**Solution:** We reflect \( F_X \) via the line that passes through \((0,0)\) and \((1,1)\) and then make it into a function.

(3) Binary hypothesis testing with continuous type observation.
Problem setting is exactly same as the discrete type observation.

- We have a system that generates continuous-type random variables.

- System is either in state \( H_0 \) or \( H_1 \).

\[
\begin{align*}
\text{if in } H_0, \text{ pdf of } X & \text{ is given by } f_0 \\
\text{if in } H_1, \text{ pdf of } X & \text{ is given by } f_1
\end{align*}
\]
. if in $H_0$, pdf of $X$ is given by $f_0$

. if in $H_1$, pdf of $X$ is given by $f_1$

We observe an outcome from the system, i.e., we observe $\{X = x\}$

There is a decision rule that assigns hypothesis to each outcome

Based on the decision rule, we decide system is in state $H_0$ or $H_1$.

Notice that $P(X=x) = 0$, since $X$ is continuous-type. However, $P(X \in (u-u+e)) \approx 2e f(x, u)$.

So, for continuous-type observations, we focus on pdf instead of pmf. Hence, we define the likelihood ratio by

$$
\Lambda(x) = \frac{f_1(x)}{f_0(x)}
$$

We are only interested in threshold policies:

(i) general threshold policies with threshold $\tau > 0$.

\[
\begin{align*}
\text{if } \Lambda(x) > \tau & \text{ then } H_1 \\
\text{if } \Lambda(x) < \tau & \text{ then } H_0 \\
\text{if } \Lambda(x) = \tau & , \text{ either}
\end{align*}
\]

(ii) ML threshold policy with $\tau = 1$

(map threshold policy with $\tau = \frac{f_0}{f_1}$, $\pi_0 = P(\text{system in } H_0), \pi_1 = P(\text{system in } H_1)$ are priors.

Instead of likelihood ratio we can consider log-likelihood:

$$
\log \Lambda(x) = \log f_1(x) - \log f_0(x)
$$

We compare it with $\log \tau$.

Note: If $u$ is not in support of $f_1$, then it means system is in $H_0$.

If $u$ is not in support of $f_0$, then it means system is in $H_1$.

We have similar notions as before:

$$
P_{\text{false alarm}} = P(\text{Decision rule is } H_1 | \text{System in } H_0)
$$

$$
P_{\text{missed}} = P(\text{Decision rule is } H_0 | \text{System in } H_1)
$$

$$
P_{\text{error}} = P(\text{Decision rule is } \text{state of system})
$$

$$
= \pi_0 P_{\text{false alarm}} + \pi_1 P_{\text{missed}}, \text{ same as in discrete-type}
$$

Example:

3.31. [A simple hypothesis testing problem with continuous-type observations]

Consider the hypothesis testing problem in which the pdf’s of the observation $X$ under hypotheses $H_0$ and $H_1$ are given, respectively, by:
3.31. [A simple hypothesis testing problem with continuous-type observations]
Consider the hypothesis testing problem in which the pdf’s of the observation $X$ under hypotheses $H_0$ and $H_1$ are given, respectively, by:

$$f_0(u) = \begin{cases} \frac{1}{2} & \text{if } -1 \leq u \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

and

$$f_1(u) = \begin{cases} |u| & \text{if } -1 \leq u \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Assume the priors on the hypotheses satisfy $\pi_1 = 2\pi_0$.

(a) Find the MAP rule.

(b) Find $p_{\text{false alarm}}$, $p_{\text{miss}}$ and the average probability of error, $p_e$, for the MAP rule.

\textbf{Solution.} Notice that the likelihood ratio is given by $L(u) = \frac{f_1(u)}{f_0(u)} = 2|u|$ for $u \in [-1,1]$.

We have to compare $L(u)$ with $\frac{\pi_1}{\pi_0} = \frac{1}{2}$ for $u \in [-1,1]$ (notice that support of $f_1$ & $f_0$ are both $[-1,1]$).

The decision rule is:

$$H_1 \text{ if } 1|u| > \frac{1}{4}$$

$$H_0 \text{ if } 1|u| < \frac{1}{4}$$

$$\text{other}\ if \ 1|u| = \frac{1}{4}$$

\begin{itemize}
  \item $X_1 = \frac{1}{2}X_0$ and $X_0 + X_1 = 1 \Rightarrow X_0 = \frac{1}{2}$ and $X_1 = \frac{1}{2}$
  \item $p_{\text{false alarm}} = P(\text{Decision rule is } H_1 \mid \text{system in } H_0)$
    \begin{align*}
      &= \int_{1|u| > \frac{1}{4}} f_0(u) \, du = \int_{\frac{1}{2} \leq u \leq \frac{1}{2} + \frac{3}{4}} \frac{1}{2} \, du = \frac{3}{4} - \frac{3}{16} \\
      &= \frac{9}{4} - \frac{3}{16} = \frac{27}{16} - \frac{3}{16} = \frac{7}{4}
    \end{align*}
  \item $p_{\text{miss}} = P(\text{Decision rule is } H_0 \mid \text{system in } H_1)$
    \begin{align*}
      &= \int_{1|u| < \frac{1}{4}} f_1(u) \, du = \int_{-\frac{1}{4}}^{\frac{1}{4}} u \, du = 2 \left( \frac{1}{2} \right) \left( \frac{1}{16} \right) = \frac{1}{16}
    \end{align*}
  \item $p_{\text{error}} = X_1 p_{\text{false alarm}} + X_0 p_{\text{miss}} = \frac{1}{3} \cdot \frac{7}{4} + \frac{2}{3} \cdot \frac{1}{16} = \frac{7}{24}
\end{itemize}