1.017/1.010 Class 8
Expectation, Functions of a Random Variable

Mean, variance of random variables

Expectation (population **mean**) of $\boldsymbol{x}$... $E[\boldsymbol{x}]$

For a **discrete** $\boldsymbol{x}$:

$$E(\boldsymbol{x}) = \bar{\boldsymbol{x}} = \sum x_i p_x(x_i)$$

For a **continuous** $\boldsymbol{x}$:

$$E(\boldsymbol{x}) = \bar{\boldsymbol{x}} = \int_{-\infty}^{+\infty} x f_x(x) dx$$

(Population) **variance** of $\boldsymbol{x}$... $Var[\boldsymbol{x}]$:

$$Var(\boldsymbol{x}) = E\left[(\boldsymbol{x} - \bar{\boldsymbol{x}})^2\right]$$

Derive these for uniform & triangular distributions

Functions of a random variable

$\boldsymbol{y} = g(\boldsymbol{x})$ $\boldsymbol{x}$ is a random variable with CDF $F_{\boldsymbol{x}}(x)$
$\boldsymbol{y}$ is a random variable with CDF $F_{\boldsymbol{y}}(y)$ since it depends on $\boldsymbol{x}$

Derived distribution problems

Derive $F_{\boldsymbol{y}}(y)$ from $F_{\boldsymbol{x}}(x)$ using either of these options:

1. Analytical derivation
   Apply definitions of $\boldsymbol{y}$, $F_{\boldsymbol{x}}(x)$, and $F_{\boldsymbol{y}}(y)$.

$$F_y(y) = P[\boldsymbol{y} \le y] = P[g(\boldsymbol{x}) \le y] = P[\boldsymbol{x} = \{x \mid g(x) \le y\}]$$

2. Stochastic simulation
   Generate many realizations of x, compute y for each replicate, construct empirical $F_{\boldsymbol{y}}(y)$ from y replicates

Example (analytical derivation): Distribution of $y = x^2$ for a uniformly distributed x:

Uniform distribution centered on 0:

$$y = g(x) = x^2$$
$$F_x(x) = \frac{x+1}{2} \quad ; \quad f_x(x) = 1 \quad ; \quad -1 \le x \le 1$$
$$F_y(y) = P[\boldsymbol{y} \le y] = P[g(\boldsymbol{x}) \le y] = P[\boldsymbol{x} = \{x \mid g(x) \le y]$$
$$F_y(y) = P[\boldsymbol{x} = \{x \mid x^2 \le y] = P[-y^{0.5} \le \boldsymbol{x} \le y^{0.5}] = F_x(y^{0.5}) - F_x(-y^{0.5}) = y^{0.5} \quad ; \quad 0 \le y \le 1$$

Uniform distribution centered on 0.5 (note need to split y interval into 2 parts):

$$y = g(x) = x^2$$
$$F_x(x) = (x+1)/3 \quad ; f_x(x) = 1/3 \quad ; \quad -1 \le x \le 2$$
$$F_y(y) = P[\boldsymbol{y} \le y] = P[g(\boldsymbol{x}) \le y] = P[\boldsymbol{x} = \{x \mid g(x) \le y]$$
$$F_y(y) = P[\boldsymbol{x} = \{x \mid x^2 \le y] = P[-y^{0.5} \le \boldsymbol{x} \le y^{0.5}] = F_x(y^{0.5}) - F_x(-y^{0.5}) = \frac{2y^{0.5}}{3} \quad ; \quad 0 \le y \le 1$$
$$= P[\boldsymbol{x} = \{x \mid x^2 \le y] = P[-1 \le \boldsymbol{x} \le y^{0.5}] = F_x(y^{0.5}) - F_x(-1) = (y^{0.5}+1)/3 \quad ; \quad 1 \le y \le 4$$

Mean and variance of $\boldsymbol{y} = g(\boldsymbol{x})$:

$$E(\boldsymbol{y}) = E[g(\boldsymbol{x})] = \sum_i g(x_i) p_x(x_i)$$

$$E(\boldsymbol{y}) = E[g(\boldsymbol{y})] = \int_{-\infty}^{+\infty} g(x) f_x(x) dx$$

$$Var(\boldsymbol{y}) = Var[g(\boldsymbol{y})] = E\left\{[g(\boldsymbol{x}) - \overline{g(\boldsymbol{x})}]^2\right\}$$