Week 5

2.1: Functions of Random Variables

RECALL

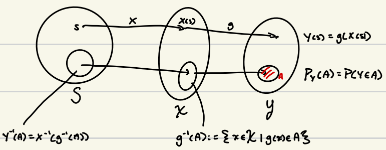
Recall (relevant ch. 1 content): A RV X is defined as a function from a sample space S to $\mathcal{X}$ (previously $\mathcal{X}$)
s.t. $X = X(s)$, $s \in S$.
Information about the RV X is (completely) captured by its cdf:

$$F_X(x) = P(X \le x) = P(s \in S \mid X(s) \le x)$$

Take a function $\underline{g: \mathcal{X} \to \mathcal{Y}}$

Then $\underline{Y = g(X)}$ is also a RV taking values in $\mathcal{Y}$ because Y is in fact a function on S, i.e.

$$Y(s) = g(X(s)) = g \circ X(s) \ne X \circ g$$



Q: what is the prob. dist. of $Y = g(X)$?

Prop. By def., $\forall A \subset \mathcal{Y}$,

$$\boxed{\begin{aligned} P(Y \in A) &= P(g(X) \in A) \\ &= P(x \in g^{-1}(A)) \quad = P(X(s) \in g^{-1}(A)) \\ &= P(s \in X^{-1}(g^{-1}(A))) = P(s \in X^{-1} \circ g^{-1}(A)) \\ &= P(s \in (g \circ X)^{-1}(A)) \\ &= P(s \in Y^{-1}(A)) \\ &= P(Y(s) \in A) \end{aligned}}$$

Note: we can 'create' new RVs by applying a function to an existing RV, instead of specifying the prob. dist. from scratch, i.e. via a sample space and sigma algebra

Ex (Binomial Transformation): X is a binomial RV if its pmf

$$f_X(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \dots, n$$

we write $X \sim \text{Binomial}(n, p)$

Let $Y := g(X)$ where $g(x) = n - x$,
i.e. $Y = n - X$ is also a RV

Q: what is the distribution of Y?

$Y \in \{0, 1, \dots, n\}$.

For $y \in \{0, 1, \dots, n\}$: $f_Y(y) := P(Y = y)$

$$= \sum_{x \in g^{-1}(y)} f_X(x)$$

$$= f_X(n - y)$$

$$= P(X = n - y) \quad , Y = n - X \Rightarrow X = n - Y$$

$$= f_X(n-y) = \binom{n}{n-y} p^{n-y} (1-p)^y$$

$$= \binom{n}{y} (1-p)^y p^{n-y} \quad \text{since } \binom{n}{y} = \frac{n!}{y!(n-y)!} = \frac{n!}{(n-y)!\,y!} = \frac{n!}{(n-y)!\,[n-(n-y)]!} = \binom{n}{n-y}$$

$$\Rightarrow Y \sim \text{Binomial}(n, 1-p).$$

NEW MATERIAL

Continuous RV's

If X is a cont. RV, g a nice (cont.) function,
then $Y = g(X)$ is a cont. RV.

so,

$$\begin{aligned} F_Y(y) &= P(Y \le y) \\ &= P(g(X) \le y) \\ &= P(x \in \mathcal{X} : g(x) \le y) \\ &= \int_{\{x : g(x) \le y\}} f_X(x)\,dx \end{aligned}$$

The set $\{x : g(x) \le y\}$ may be difficult to identify so F_Y may be hard to derive in general.

Prop. If g is (strictly) monotone, this gets simpler s.t. either
$$\left[\begin{array}{l}\cdot\ g \text{ is increasing, i.e. } g(x)>g(y) \text{ if } x>y\\ \cdot\ g \text{ is decreasing, i.e. } g(x)<g(y) \text{ if } x>y\end{array}\right.$$

recall: monotonic function values only move in one direction (increasing/decreasing) as the input increases

note: $g^{-1}(y)=\{x : g(x)=y\}$ is a singleton set.

Thm. Let X have the cdf $F_X(x)$,

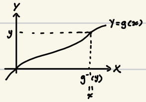
$\mathcal{X}=\{x : f_X(x)>0\} \leftarrow$ support of $\{ {}^{X}_{f_X}$

$\mathcal{Y}=g(\mathcal{X}) := \{y : y=f(x) \text{ for some } x\in\mathcal{X}\}$

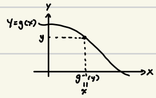
a) If $g:\mathcal{X}\to\mathcal{Y}$ is (strictly) increasing, then $Y=g(X)$ is a RV taking values values in $\mathcal{Y}$ with the cdf $\boxed{F_Y(y)=F_X(g^{-1}(y))\ \forall y\in\mathcal{Y}}$

b) If $g:\mathcal{X}\to\mathcal{Y}$ is (strictly) decreasing, then $Y=g(X)$ is a RV in $\mathcal{Y}$ with cdf $\boxed{F_Y(y)=1-F_X(g^{-1}(y))\ \forall y\in\mathcal{Y}}$

Proof (a). If $g\uparrow$ (strictly increasing), where $Y=g(X)$, then
$\{x: g(x)\le y\}=\{x\le g^{-1}(y)\}$
$$\Rightarrow F_Y(y)=\int_{\{x\le g^{-1}(y)\}} f_X(x)\,dx = F_X(g^{-1}(y)).\ \blacksquare$$



Proof (b). If $g\downarrow$ (strictly decreasing), where $Y=g(X)$, then
$\{x: g(x)\le y\}=\{x\ge g^{-1}(y)\}$
$$\Rightarrow F_Y(y)=\int_{\{x\ge g^{-1}(y)\}} f_X(x)\,dx = 1-F_X(g^{-1}(y)).\ \blacksquare$$



corollary: we can deduce the pdf from cdf:

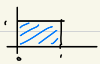
from (a), $f_Y(y)=\frac{d}{dy}F_Y(y)$
$$= f_X(g^{-1}(y))\underbrace{\frac{d}{dy}g^{-1}(y)}_{>0 \text{ since } g\uparrow\Rightarrow g^{-1}\uparrow}.$$

and from (b),
$$f_Y(y)=-f_X(g^{-1}(y))\underbrace{\frac{d}{dy}g^{-1}(y)}_{<0 \text{ since } g\downarrow\Rightarrow g^{-1}\downarrow}.$$

combining the two cases to obtain

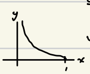
$$\boxed{\underline{\text{change-of-var (formula)}}:=\ f_Y(y)=f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|}$$

Ex (uniform & exponential): Let $X\sim \text{Uniform}(0,1)$, i.e. $f_X(x)=\begin{cases}1, & x\in(0,1)\\ 0, & \text{otherwise}\end{cases}$



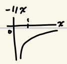
Take $Y=-\log(X)$.

Q: what is the dist. of Y?



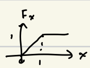
$Y=g(X)$ where $g(x)=-\log(x)$ for $\begin{cases}\mathcal{X}=(0,1)\\ \mathcal{Y}=(0,+\infty)\end{cases} \leftarrow$ support

Solve x for $y=-\log(x) \Rightarrow \log(x)=-y \Rightarrow x=e^{-y}$.



$g'(x)=\frac{-1}{x}<0 \Rightarrow g\downarrow$ and $g^{-1}(y)=e^{-y}\ \forall y>0$

If $x\in(0,1)$, then $F_X(x)=\int_0^x f_X(x)\,dx = x$



Thus, for $y>0$
$$F_Y(y)=1-F_X(g^{-1}(y))=1-g^{-1}(y)=1-e^{-y}$$

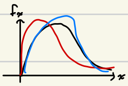
Hence $f_Y(y)=\frac{dF_Y(y)}{dy}=(-e^{-y})'\frac{d}{dy}(-y)=-e^{-y}(-1)=e^{-y}$ for $y>0$.

For $y\le 0$, $F_Y(y)=0 \Rightarrow f_Y(y)=0$ as well

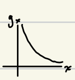
we say $Y\sim \text{Exp}(1)$.

Recall: Exponential(θ) contains pdf $:=\frac{1}{\theta}\exp\left(-\frac{x}{\theta}\right)$; $x\ge 0, \theta>0$.

Ex (inverse Gamma): Let $X\sim \text{Gamma}(n,\beta)$ where n,β: shape and scale parameters respectively



Suppose $Y=\frac{1}{X}$, i.e. the inverse gamma RV where
$$f_X(x)=\frac{1}{(n-1)!\beta^n}x^{n-1}e^{-x/\beta} \text{ for } x>0\ (0,\infty)$$

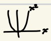


Given $Y=g(X)=\frac{1}{X} \Rightarrow x=g^{-1}(y)=\frac{1}{y}$ for $y\in(0,\infty)=\mathcal{Y}$.

$g\downarrow$ on $\mathcal{X}=(0,\infty)$ and $\frac{d}{dy}[g^{-1}(y)]=\frac{d}{dy}(y^{-1})=\frac{-1}{y^2}$

we know
$$f_Y(y)=f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|$$
$$=\frac{1}{(n-1)!\beta^n}(1/y)^{n-1}e^{-1/(\beta y)}\frac{1}{y^2}$$
$$=\frac{1}{(n-1)!\beta^n}y^{-(n+1)}e^{-1/(\beta y)}$$

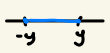
__Ex (Square Transformation)__: Suppose $X \sim Normal(0,1)$ and let $Y = g(X) = X^2 \sim \chi^2_1$.

Here, $g(x) = x^2$ for $x \in \mathbb{R}$ is $\underline{\underline{\text{not}}}$ a monotonic function 

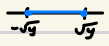
Since $\mathcal{X} = \mathbb{R}$, $\mathcal{Y} = \mathbb{R}_+ \cup \{0\}$.

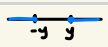
Thus,

$\forall y \geq 0$, $F_Y(y) = P(Y \leq y) = P(X^2 \leq y)$

$|x| \leq y = \begin{cases} x \leq y \\ x \geq -y \end{cases}$ 

$= P(|X| \leq \sqrt{y})$, $x \in \mathbb{R}$

$= P(X \leq \sqrt{y}) - P(X \leq -\sqrt{y})$ 

$|x| \geq y = \begin{cases} x \geq y \\ x \leq -y \end{cases}$ 

$= F_X(\sqrt{y}) - F_X(-\sqrt{y})$

So,

$$f_Y(y) = \frac{d}{dy}[F_Y(y)] = \frac{d}{dy}\left[F_X(y^{1/2}) - F_X(-y^{1/2})\right] = \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(-\sqrt{y})$$

$$= \frac{1}{2\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-y/2} + \frac{1}{2\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-y/2}$$

since $X \sim N(0,1) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, $x \in \mathbb{R}$ and $\chi^2_{df} \overset{d}{\sim} \sum_{i=1}^{df} z_i^2$ for $z_i \overset{iid}{\sim} N(0,1)$.

$$= \frac{1}{\sqrt{2\pi y}} e^{-y/2} \sim \chi^2_1$$

$$\propto y^{\frac{1}{2}-1} e^{-y/2} = Gamma\left(\tfrac{1}{2}, 2\right)$$

__Thm (Probability Integral Transform)__. Let X - continuous RV w/ cdf $F_X(x)$.
(1)
Let $Y = F_X(x) \iff g(x) = F_X(x)$
Then
$Y \sim Uniform(0,1)$

__Remarks__:
- Doesn't hold for discrete RV's (F_X would be a step function)
- F_X might not be strictly increasing from 0 to 1
- Define the __inverse function__ of F_X as

$$\boxed{F_X^{-1}(y) := \inf\{x : F_X(x) \geq y\} \ \forall y \in (0,1)}$$

where
- $F_X^{-1}(1) := +\infty$ only if $F_X(x) < 1 \ \forall x$.
- $F_X^{-1}(0) := -\infty$

__Recall__: Let $A \subset \mathbb{R}$.

then $\inf A = \begin{cases} \text{largest } a \in \mathbb{R} \cup \{-\infty\} \text{ s.t. } a \leq x \ \forall x \in A. \\ +\infty \text{ if } a \text{ doesn't exist} \end{cases}$

where $\inf([a,b]) = a$

and $\inf((a,b]) = a \neq \min(a,b]$ which DNE

__Proof (Sketch).__ I will prove the following theorem:

(1)
Let X - continuous RV w/ cdf $F_X(x)$ and $Y = g(x) = F_X(x)$.
Then, $Y \sim Uniform(0,1)$

__Proof.__ For $y \in (0,1)$,

$P(Y \leq y) = P(F_X(x) \leq y)$

assuming $F_X^{-1} \uparrow$ (i.e. strictly increasing)

$= P(F_X^{-1}(F_X(x)) \leq F_X^{-1}(y))$

$= P(X \leq F_X^{-1}(y))$ (*) requires further discussion

$= F_X(F_X^{-1}(y))$

assuming F_X is continuous

$= y$.

It is easy to check: $P(Y \leq 1) = 1$ and $P(Y \leq 0) = 0$.

Hence, cdf of Y is that of a unif(0,1) RV. □

(*):

- If F_X is strictly increasing at x, for $x \in \mathcal{X}$

  then let $y = F_X(x) \Rightarrow x = F_X^{-1}(y)$

  so $F_X^{-1}(F_X(x)) = F_X^{-1}(y) = x$.

- If F_X is "flat" at x, i.e. $\forall x \in$ some $[x_1, x_2]$

  s.t. $F_X(x') = F_X(x_1)$ $\forall x_1 \in [x_1, x]$

  and $F_X(x) < F_X(x_1)$ $\forall x < x_1$

  then

  $F_X^{-1}(F_X(x)) = x_1 (\leq x)$

  since

  $P(X \in [x_1, x]) = 0$

  $P(X \in (F_X^{-1}(F_X(x)), x]) = 0$ (**)

  So in this scenario

  $\forall a \in \mathbb{R}$

  $P(F^{-1}(F_X(X)) \leq a \mid F_X \text{ is flat at } X)$

  $= P(F_X^{-1}(F_X(X)) \leq a, X \leq a \mid \cdots) +$

  $P(F_X^{-1}(F_X(X)) \leq a, X > a \mid \cdots)$

  $=: C + D$, where

  $C = P(X \leq a \mid F_X \text{ is flat at } X)$

  b/c $X \leq a \Rightarrow F_X^{-1}(F_X(X)) \leq X \leq a$

  and $D \leq P(F_X^{-1}(F_X(X)) < X \mid F_X \text{ is flat at } X)$

  $= P(X \in (F_X^{-1}(F_X(X)), X] \mid F_X \text{ is flat at } X)$

  with prob. 1

  $= 0$ due to (**)

combining these two scenarios:

$\forall a$, $P(F_X^{-1}(F_X(X)) \leq a \mid F_X \text{ flat at } X) = P(X \leq a \mid F_X \text{ flat at } X)$

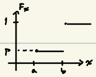
$P(F_X^{-1}(F_X(X)) \leq a \mid F_X \uparrow \text{ at } X) \underset{\text{w/ prob 1}}{=} P(X \leq a \mid F_X \uparrow \text{ at } X)$

to qualify (*). □

__Thm.__ Let $F_X(x)$ be a cdf of a RV.
(2)
Let $Y \sim Uniform(0,1)$
and set $Z = F_X^{-1}(Y)$.
Then Z has cdf F_X

__Remarks__
- useful for generating RV's from Uniform RV's
- No restriction to continuity; holds generally (for real-valued RV's)

__Proof (2).__ Simple case $X \in \{a, b\}$ where $P(X=a) = p$, $P(X=b) = 1-p$, $a < b$ and $Y \sim$ uniform.



$$F_X^{-1}(y) = \begin{cases} -\infty, & y = 0 \\ a, & 0 < y \leq p \\ b, & p < y \leq 1 \end{cases}$$

see that $P(F_X^{-1}(y) = a) = P(0 < Y \leq p) = F_Y(p) = p$ since $Y \sim$ uniform

$P(F_X^{-1}(y) = b) = P(p < Y \leq 1) = 1 - p$

$P(F_X^{-1}(y) = -\infty) = P(Y = 0) = 0$

we conclude that $F_X^{-1}(y) \overset{d}{=} X$. ■

2.2: Expectation

Def: The <u>expectation</u> of a RV $g(X)$ is

$$E[g(X)] := \begin{cases} \int_{-\infty}^{\infty} g(x) f_X(x)\,dx, & X\text{-continuous} \\ \sum_{x \in \mathcal{X}} g(x) \underbrace{f_X(x)}_{P(X=x)}, & X\text{-discrete} \end{cases} \quad (*)$$

<u>Remarks</u>: • Also known as 'expected values', 'average' of a RV or of the prob. dist. of the RV $g(X)$.

Q: what is the expectation of X?

If $X \in \mathbb{R}$, then by letting $g(X) := x$:

$$E(X) = \begin{cases} \int x f(x)\,dx, & X\text{-cont.} \\ \sum_{x \in \mathcal{X}} x \cdot f_X(x), & X\text{-discrete} \end{cases} \quad (**)$$

<u>Note</u>: If the domain of $\mathcal{X}$ is not a subset of $\mathbb{R}$ (Euclidean space), then $(**)$ may be invalid but a notation of $E(X)$ may still be designed via the expectations of $\{E g(X)\}$ in $(*)$

<u>Ex (Exponential)</u>. Let $X \sim Exp(\lambda) = \frac{1}{\lambda} e^{-x/\lambda}$, $x \in \mathcal{X} = (0, +\infty)$, $\lambda > 0$.

We can show that $\int_x f_X(x)\,dx = \int_0^\infty \frac{1}{\lambda} e^{-x/\lambda}\,dx = \frac{1}{\lambda}[e^{-x/\lambda}(-\lambda)]\Big|_0^\infty = -e^{-x/\lambda}\Big|_0^\infty = 0-(-1) = 1.$

$$E(X) = \int_0^\infty x\left(\frac{1}{\lambda} e^{-x/\lambda}\right)dx$$

↳ IBP: $\int u\,dv = uv - \int v\,du$ L I A T E

let $u = x$, $dv = \frac{1}{\lambda} e^{-x/\lambda}\,dx$

$du = 1$, $v = \int \frac{1}{\lambda} e^{-x/\lambda}\,dx = e^{-x/\lambda} \frac{1}{\lambda}(-\lambda) = -e^{-x/\lambda}$

$$= -x e^{-x/\lambda}\Big|_0^\infty + \int_0^\infty e^{-x/\lambda}\,dx$$

$$= 0 + e^{-x/\lambda}(-\lambda)\Big|_0^\infty$$

$$= 0 - (-\lambda) = \lambda.$$

<u>Ex (Binomial Expectation)</u>: Let $X \sim Binomial(n,p)$

where $f_X(x) = P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$, $x \in \mathcal{X} = \{0, \dots, n\}$, $n \in \mathbb{N}$, $p \in (0,1)$.

First check that

$$\sum_{x \in \mathcal{X}} f_X(x) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = (p + (1-p))^n \text{ by the Binomial identity: } (x+a)^n = \sum_{k=0}^{n} \binom{n}{k} x^k a^{n-k}$$

$$= 1.$$

Then,

$$E(X) = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum_{x=1}^{n} x \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0 \Rightarrow \text{entire product}.$$

$$= \sum_{x=1}^{n} n \binom{n-1}{x-1} p^x (1-p)^{n-x}, \quad x\binom{n}{x} = x \cdot \frac{n!}{x!(n-x)!} = \frac{n!}{(x-1)!(n-x)!} = n \cdot \frac{(n-1)!}{(x-1)![(n-1)-(x-1)]!} = n\binom{n-1}{x-1}.$$

$$= np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} (1-p)^{n-x}$$

letting $y = x-1 \Rightarrow -x = -1-y$

$$= np \sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{n-1-y}$$

$$= np(p + \cancel{(1-p)})^{n-1}$$

$$= np.$$