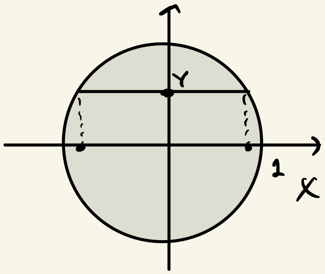
Conditional distributions



- Intuitive examples:
  - $X|Y \sim N(0,Y^2)$ where $Y \sim \text{Unif}[0,1]$
  - $(X,Y) \sim \text{Unif}(\text{unit disc}) \Rightarrow X|Y \sim \text{Unif}\left(-\sqrt{1-Y^2}, \sqrt{1-Y^2}\right)$
- General def:

Def Let X be a r.v. and $\mathcal{F} \subset \Sigma$ be a σ-algebra. Consider

$$\mu(B,\omega) := \mathbb{P}\{X \in B \mid \mathcal{F}\}_{(\omega)} = \mathbb{E}\left[\mathbb{1}_{\{X\in B\}} \mid \mathcal{F}\right](\omega) \qquad \forall B \in \mathcal{B}.$$

If $\mu(\cdot,\omega)$ is a prob. measure on $\mathcal{B}(\mathbb{R})$ for all $\omega \in \Omega$, ← or all ω except a set of $\mathbb{P}$ zero.
we call it the conditional distribution of X on $\mathcal{F}$.
↖random measure

A r.v. with distribution $\mu(B,\omega)$ is denoted $X|\mathcal{F}$ ($X|Y$ if $\mathcal{F} = \sigma(Y)$).

Prop (Conditional density) If (X,Y) has density $f_{X,Y}(x,y)$, then $X|Y$ has density

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, \quad x \in \mathbb{R}, \qquad \text{whenever } Y = y.$$

↖(marginal) density $= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx$

Proof • $f_{X|Y}(\cdot|y)$ is indeed a density $\forall y$ (total integral = 1)

• It remains to check that $\forall B \in \mathcal{B}(\mathbb{R})$

$$\int_B f_{X|Y}(x,Y)\,dx \overset{?}{=} \mathbb{E}\left[\mathbb{1}_{\{X\in B\}} \mid \sigma(Y)\right] \quad \text{a.s. (in } Y)$$

def of cond.exp. $\Longleftrightarrow$

$$\mathbb{E}\left[\left(\int_B f_{X|Y}(x,Y)\,dx\right)\mathbb{1}_F\right] \overset{?}{=} \mathbb{E}[\mathbb{1}_{\{X\in B\}}\mathbb{1}_F\} \quad \forall\, \overset{\{Y\in R\}\text{ for some } R\in\mathcal{B}(\mathbb{R})}{\overset{\shortparallel}{F}} \in \sigma(Y)$$

$$\int_R\left(\int_B f_{X|Y}(x,y)\,dx\right) f_Y(y)\,dy \overset{?}{=} P\{X\in B, Y\in R\}$$

$$= \int_{R\times B} f(x,y)\,dx\,dy.$$ Yes! D.

Remark | $f_{X|Y} = \frac{f_{X,Y}}{f_Y}$ is a "continuous version" of $P(E|F) = \frac{P(E\cap F)}{P(F)}$

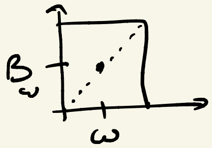
Allows to condition on events $\{Y=y\}$ of measure 0 ☺

"Continuous" version of LTP: $f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y)\,dy$; Bayes: $f_{Y|X} = \frac{f_{X|Y}\cdot f_Y}{f_X}$, etc.

REGULAR CONDITIONAL DISTRIBUTIONS (SKIPPED)

- Unfortunately, $\mu(\cdot,\omega)$ in Def p-153 may not be a prob. meas a.s. ☹

(idea of an example: destroy the LHS for set $B=B_\omega$ $\forall \omega\in\Omega$)
$\Rightarrow$ LHS is NOT a prob. measure $\forall\omega$.



- Nevertheless:

THM (Conditional distributions) let X be a r.v. and $\mathcal{F}\subset\Sigma$ be a σ-algebra.
Then $\forall B\in\mathcal{B}(\mathbb{R}), \forall\omega\in\Omega$ $\exists\ \mu(B,\omega)$ such that:

(i) $\forall\omega\in\Omega$, $\mu(\cdot,\omega)$ is a probability measure on $\mathcal{B}(\mathbb{R})$

(ii) $\forall B\in\mathcal{B}(\mathbb{R})$, $\mu(B,\cdot)$ is a version of $\mathbb{E}[\mathbb{1}_{\{X\in B\}}|\mathcal{F}]$
i.e. agree a.s.

$\mu(\cdot,\cdot)$ is called a Markov kernel.

Proof. $\forall r\in\mathbb{Q}$, consider

$$F(r,\omega):=\forall\text{ version of }\mathbb{E}[\mathbb{1}_{\{X\le r\}}|\mathcal{F}](\omega) \qquad (\circ)$$

Then $\exists$ "nice" set $A\subset\Omega$ with $\mathbb{P}(A)=1$ such that $\forall\omega\in A$ we have:

$(*)$ $F(r,\omega)<F(s,\omega)$ $\forall\ r<s$ in $\mathbb{Q}$ ($A\subset B\Rightarrow\mathbb{E}[\mathbb{1}_A|\mathcal{F}]\le\mathbb{E}[\mathbb{1}_B|\mathcal{F}]$ a.s.)

$(**)$ $F(r+\frac{1}{n},\omega)\to F(r,\omega)$ $\forall r\in\mathbb{Q}$ (conditional monotone convergence thm)

$(***)$ $F(n,\omega)\to\infty$, $F(-n,\omega)\to 0$ (same)

- For $\omega\in A$, extend $(\circ)$ to $\mathbb{R}$ by defining

$$F(x,\omega):=\inf\{F(r,\omega): r>x\} \qquad \forall x\in\mathbb{R}. \qquad (\infty)$$

For $\omega\notin A$, let F be some fixed cdf, e.g. $F(x,\omega):=\Phi(x)$ $\forall x\in\mathbb{R}$.
$N(0,1)$

SKIPPED

- $(*), (**) \Rightarrow F(x,\omega)$ in (oo) agrees with (o) on all rational x, is right-continuous, monotone and has correct limits $(0/1)$

  $\Rightarrow F(\cdot,\omega)$ is a CDF $\forall \omega \in \Omega$.

  $\Rightarrow \exists$ prob. measure $\mu(\cdot,\omega)$ with cdf $F(\cdot,\omega)$.

- It remains to prove (ii), i.e. to show that $\forall B \in \mathcal{B}(\mathbb{R})$:

  (a) $\mu(B,\cdot)$ is $\mathcal{F}$-measurable

  (b) $\int_F \mu(B,\omega)\,dP(\omega) = \mathbb{E}\left[\mathbb{1}_{\{X\in B\}}\mathbb{1}_F\right] \quad \forall F \in \mathcal{F}. \qquad (\bigstar)$

(a) The family $\mathcal{P} := \{B \in \mathcal{B} : \mu(B,\cdot)$ is $\mathcal{F}$-measurable$\}$ is a Λ-system

$$\left(\mu\Big(\bigsqcup_{i=1}^{\infty} B_i,\omega\Big) \overset{\mu(\cdot,\omega)\text{ is a prob. measure}}{=} \sum_{i=1}^{\infty}\mu(B_i,\omega). \text{ If each function in RHS is mble} \Rightarrow \text{their sum is mble} \right)$$

and contains all intervals $(-\infty, r]$, $r \in \mathbb{Q}$ by (o)

Dynkin's π-λ theorem $\Rightarrow \mathcal{P} \supset \sigma\big((-\infty,r] : r\in\mathbb{Q}\big) = \mathcal{B}(\mathbb{R})$.

(b) Fix $\forall F \in \mathcal{F}$.

Each side of $(\bigstar)$ defines a prob. measure as a function of B, and the equation holds for $\forall B = (-\infty, r]$, $r \in \mathbb{Q}$

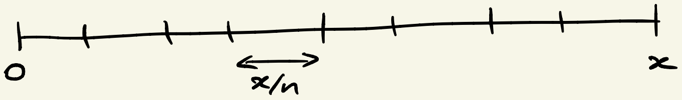
$$\left(\text{LHS} = \int_F F(r,\omega)\,dP(\omega) \overset{(o)}{=} \mathbb{E}\left[\mathbb{1}_{\{X\le r\}}\mathbb{1}_F\right] \right)$$

$\Rightarrow$ by the uniqueness of measures, $(\bigstar)$ holds $\forall B \in \mathcal{B}(\mathbb{R})$. $\square$

EXPONENTIAL DISTRIBUTION

Ex | X = time of the first call at a police station after midnight.
Distribution of X?

cdf $F(a) = P\{X \le x\} = ?$



Divide $[0,x)$ into n intervals of length a/n.

$P\{\text{call during a given interval}\} = \lambda \cdot \frac{x}{n}$ (proportional to the length of the interval)

coefficient of proportionality

$\Rightarrow P\{X > a\} = P\{\text{no call in } [0,x)\}$

$= P\{\text{no call in any of the intervals}\}$

$$= \left(1 - \frac{\lambda x}{n}\right)^n \longrightarrow e^{-\lambda x} \quad \text{as } n \to \infty$$

Thus, $P\{X > x\} = e^{-\lambda x}, \quad a > 0$

$\Downarrow$

Def A r.v. X has the exponential distribution with parameter λ if

$$P\{X > x\} = e^{-\lambda x}, \quad x > 0.$$

Notation: $X \sim \text{Exp}(\lambda)$. The parameter λ is called the rate.

- pdf: $f(x) = \frac{d}{dx}(1 - e^{-\lambda x}) = \begin{cases} \lambda e^{-\lambda x}, & x \ge 0 \\ 0, & x < 0 \end{cases}$
- $EX = \int_0^\infty x e^{-\lambda x} dx = \boxed{\frac{1}{\lambda}}$, $\quad Var(X) = \boxed{\frac{1}{\lambda^2}}$ (DIY)
- Exponential distribution is used to model waiting times (lifetime of iPhone, time until next customer arrives, etc.)

Prop (Memoryless property) $X \sim \mathrm{Exp}(\lambda)$ satisfies

$$\underset{\substack{\uparrow \\ P\{\text{wait} > s \text{ more minutes}\}}}{P\{X > t+s \mid X > t\}} = \underset{\substack{\uparrow \\ P\{\text{wait} > s \text{ minutes}\}}}{P\{X > s\}} \qquad \forall s,t>0$$

$$LHS = \frac{P\{X>t+s\}}{P\{X>t\}} \underset{\substack{\uparrow \\ \text{def}}}{=} \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} = RHS$$

- i.e. $X \sim \mathrm{Exp}(\lambda) \Rightarrow X - t \mid X \geq t \sim \mathrm{Exp}(\lambda)$

- **Ex①** let $X_i \sim \mathrm{Exp}(\lambda_i)$, $i=1,2$, be independent.

(a) $$P\{X_1 \leq X_2\} = \mathbb{E}_{X_1}[X_1 \leq X_2 \mid X_1] \overset{\substack{\text{tail of } X_2 \\ \downarrow}}{=} \mathbb{E}_{X_1}\left[e^{-\lambda_2 X_1}\right] = \int_0^\infty e^{-\lambda_2 x} \cdot \overbrace{\lambda_1 e^{-\lambda_1 x}}^{\text{pdf of } X_1} dx = \frac{\lambda_1}{\lambda_1+\lambda_2}$$

(b) $\min(X_1, X_2) \sim \mathrm{Exp}(\lambda_1+\lambda_2)$

$$P\{\min(X_1,X_2) > t\} = P\{X_1 \geq t, X_2 \geq t\} \overset{\substack{\text{indep} \\ \downarrow}}{=} e^{-\lambda_1 t} e^{-\lambda_2 t} = e^{-(\lambda_1+\lambda_2)t}$$

induction ⇒ **Cor** $\min(X_1, \dots, X_n) \sim \mathrm{Exp}(\lambda_1 + \dots + \lambda_n)$ (Min-stability)

Ex② You arrive at a post office having 2 clerks; both are busy; no line. You enter the service when either clerk becomes free. The service times of clerks are $\mathrm{Exp}(\lambda_1)$, $\mathrm{Exp}(\lambda_2)$. Find the expected $\underbrace{\text{time you spend in the office}}_{=: T}$.

$R_i :=$ remaining service time of the customer with clerk i $\sim \mathrm{Exp}(\lambda_i)$ by memoryless property.

$S :=$ your service time. $T = \min(R_1, R_2) + S \Rightarrow$

$$\mathbb{E}T = \mathbb{E}\underbrace{\min(R_1,R_2)}_{\sim \mathrm{Exp}(\lambda_1+\lambda_2) \ (\text{Ex } 1b)} + \mathbb{E}S = \frac{1}{\lambda_1+\lambda_2} + \mathbb{E}S.$$

$$\mathbb{E}S = \underbrace{\mathbb{E}[S \mid R_1 < R_2]}_{\underbrace{\sim \mathrm{Exp}(\lambda_1)}_{\sim 1/\lambda_1}} \underbrace{P\{R_1 < R_2\}}_{= \frac{\lambda_1}{\lambda_1+\lambda_2} \ (\text{Ex } 1a)} + \underbrace{\mathbb{E}[S \mid R_2 \leq R_1]}_{\underbrace{\sim \mathrm{Exp}(\lambda_1)}_{\sim 1/\lambda_1}} \underbrace{P\{R_2 \leq R_1\}}_{= \frac{\lambda_2}{\lambda_1+\lambda_2} \ (\text{Ex. } 1a)} = \frac{2}{\lambda_1+\lambda_2} \Rightarrow \mathbb{E}T = \frac{3}{\lambda_1+\lambda_2}$$

Ex (Geometric distribution) $X \sim \text{Geom}(p)$

$\# \; X = \#$ of the first success in independent trials (each = success with prob. p).

- $$\mathbb{E}X = \overset{S}{1 \cdot p} + \overset{FS}{2(1-p)p} + \overset{FFS}{3(1-p)^2 p} + \overset{FFFS}{4(1-p)^3 p} + \dots = \frac{1}{p} \qquad \text{(check!)}$$

- Alternative computation:

  Condition on 1st trial, use L.T.E:

$$E[X] = \underbrace{E[X|S]}_{1} \cdot \underbrace{P(S)}_{p} + E[\underbrace{X|F}_{\uparrow}] \cdot \underbrace{P(F)}_{1-p}$$

$X|F = X+1$ in distribution:

after 1 failure, the experiment resets.

$\Rightarrow E[X|F] = 1 + E[X]$

$\Rightarrow$ $E[X] = p + (1 + E[X])(1-p)$. Solving yields

$$\boxed{E[X] = \frac{1}{p}}$$

Also memoryless (Check!)

Ex The only memoryless r.v. with continuous distr. is $\text{Exp}(\cdot)$

The only memoryless r.v. with discrete distr. is $\text{Geom}(\cdot)$

HW

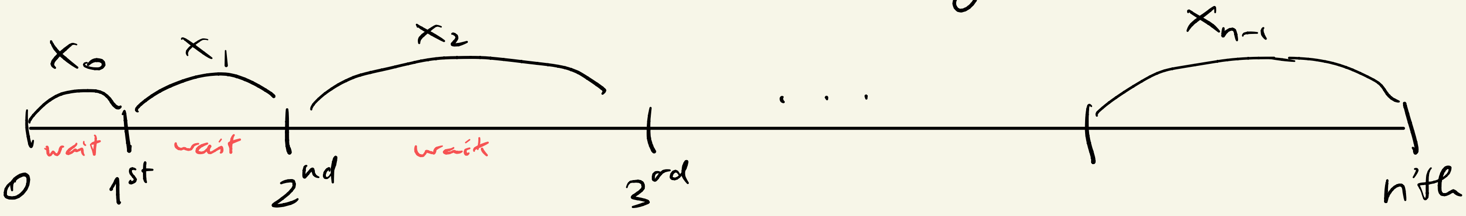
Ex (Coupon collector's problem).

> What is the expected number of coupons one needs to collect before obtaining a complete collection of all n types of coupons?

(Assume: each time one obtains a coupon, it is equally likely to be one of n types)

- $X = X_0 + X_1 + \cdots + X_{n-1}$ where

X_i = # additional coupons (after i types have been collected) in order to obtain a new type



$$E[X] = E(X_0) + E(X_1) + \cdots + E(X_{n-1})$$

- $X_i \sim ?$ After i coupons have been collected, each coupon we obtain is of a new type with probability

$$p_i = \frac{n-i}{n}$$

← # new types
← # all types

Hence $X_i \sim \mathrm{Geom}(p_i) \Rightarrow E[X_i] = \frac{1}{p_i}$

- $E[X] = \frac{1}{p_0} + \frac{1}{p_1} + \cdots + \frac{1}{p_{n-1}} = \frac{n}{n} + \frac{n}{n-1} + \cdots + \frac{n}{1} = \boxed{n\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right)}$

- Asymptotic analysis: $\sum_{k=1}^{n} \frac{1}{k} \approx \int_1^n \frac{dx}{x} = \ln(x)\Big|_1^n = \ln(n)$

$\Rightarrow \boxed{E[X] \approx n \ln n}$ Logarithmic oversampling.