Ex (Filtering)

Let $X \sim N(0,1)$ be an unknown signal, corrupted by unknown indep. noise $W \sim N(0,\sigma^2)$. You observe $Y = X + W$. Estimate X from Y, σ to minimize the mean-squared error.

- $\hat{X} := \operatorname{argmin}\{\mathbb{E}(\hat{X}-X)^2 : \hat{X} \text{ is } \sigma(Y)\text{-mble}\}$

$$\overset{\text{Cor p.151}}{=} \mathbb{E}[X|Y]$$

$$\left(\mathbb{E}X^2 = 1,\ \mathbb{E}XW = 0 \text{ (indep)},\quad X+W \sim N(0, 1+\sigma^2) \right)$$

$$\overset{\text{Remark prev. page}}{=} \frac{\mathbb{E}[XY]}{\mathbb{E}[Y^2]} \cdot Y = \frac{\mathbb{E}X(X+W)}{\mathbb{E}(X+W)^2} \cdot Y = \frac{Y}{1+\sigma^2}$$

- $$\text{MSE} = \mathbb{E}(\hat{X}-X)^2 = \mathbb{E}\left(\frac{X+W}{1+\sigma^2} - X\right)^2 = \mathbb{E}\Big(\frac{\overbrace{W-\sigma^2 X}^{N(0,\sigma^2+\sigma^4)}}{1+\sigma^2}\Big)^2 = \frac{\sigma^2+\sigma^4}{(1+\sigma^2)^2} = \frac{\sigma^2}{1+\sigma^2}$$

- **Interpretation:**
  - Small noise $(\sigma \to 0) \Rightarrow \hat{X} \approx Y$, $\text{MSE} \to 0$ (trust observation)
  - Large noise $(\sigma \to \infty) \Rightarrow \hat{X} = 0$, $\text{MSE} \to 1$ (ignore observation)

- **Normality is essential** ($X \sim$ Rademacher, $W \sim \text{Unif}(-1,1)$ $\Rightarrow$ $\hat{X} := \operatorname{sign}(X+W) = X$ is exact recovery)

HW extend to n noisy observations

$$Y_i \sim X + W_i$$

Show that $\hat{X} = \dfrac{\bar{Y}}{1+\bar{\sigma}^2}$ where $\bar{Y} = \frac{1}{n}\sum_1^n Y_i$, $\bar{\sigma}^2 = \frac{1}{n}\sigma_i^2$

Ex (Predicting a term from a sum)

Let $X_1, \dots, X_n$ be iid r.v's, $S_n := X_1 + \cdots + X_n$. Then
$$\mathbb{E}[X_i | S_n] = \frac{S_n}{n} \qquad \forall i \le n$$

$\ulcorner$ $\mathbb{E}[X_i | S_n] =: f_i(S_n)$ don't depend on i by symmetry,
and $\sum_1^n f_i(S_n) = \mathbb{E}[S_n | S_n] = S_n \Rightarrow f_i = \frac{S_n}{n}$. $\lrcorner$

Ex (Predicting a term from many sums)

Ex $\mathcal{F}_n := \sigma(S_n, S_{n+1}, \dots) \Rightarrow \mathbb{E}[X_i | \mathcal{F}_n] = \frac{S_n}{n} \quad \forall i \le n$

$\ulcorner$
$$\mathcal{F}_n \overset{\uparrow}{=} \sigma(S_n, X_{n+1}, X_{n+2}, \dots) = \sigma\Big(\underbrace{\sigma(S_n)}_{\mathcal{G}} \sqcup \underbrace{\sigma(X_{n+1}, \dots)}_{\mathcal{F}}\Big)$$

("$\subset$" from $S_k = S_n + X_{n+1} + \cdots + X_k \quad \forall k \ge n$
"$\supset$" from $X_k = S_k - S_{k-1} \quad \forall k > n$)

$$\Rightarrow \mathbb{E}[X_i | \mathcal{F}_n] \overset{p.149(c')}{=\!=\!=} \mathbb{E}[X_i | \mathcal{G}] \overset{\text{Ex. above}}{=} \frac{S_n}{n}$$
$\lrcorner$

THE LAW OF TOTAL VARIANCE

Def The <u>conditional variance</u> of a r.v. X on a σ-algebra $\mathcal{F}$ is

$$\mathrm{Var}(X\mid\mathcal{F}) := \mathbb{E}\left[\left(X-\mathbb{E}[X\mid\mathcal{F}]\right)^2 \mid \mathcal{F}\right] \overset{\text{check!}}{=} \mathbb{E}[X^2\mid\mathcal{F}] - \left(\mathbb{E}[X\mid\mathcal{F}]\right)^2 \qquad (*)$$

Prop (<u>The Law of Total Variance</u>)

$$\mathrm{Var}(X) = \mathbb{E}[\mathrm{Var}(X\mid\mathcal{F})] + \mathrm{Var}\left(\mathbb{E}[X\mid\mathcal{F}]\right)$$

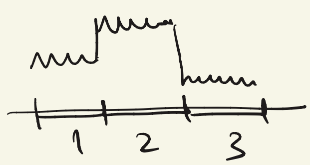
- $\mathbb{E}[X^2] \overset{\text{LTP}}{=} \mathbb{E}\left[\mathbb{E}[X^2\mid\mathcal{F}]\right] \overset{(*)}{=} \mathbb{E}\left[\mathrm{Var}(X\mid\mathcal{F}) + (\underbrace{\mathbb{E}[X\mid\mathcal{F}]}_{Z})^2\right]$
- $(\mathbb{E}X)^2 \overset{\text{LTP}}{=} (\mathbb{E}Z)^2$.

Subtract $\Rightarrow$ $\mathrm{Var}(X) = \mathbb{E}[\mathrm{Var}(X\mid\mathcal{F})] + \underbrace{\mathbb{E}[Z^2] - (\mathbb{E}Z)^2}_{\mathrm{Var}(Z)}$ $\lrcorner$

Ex: Sample a random student in a calculus class

X = student's exam score

Y = student's group number (e.g. 1, 2, 3)



$\Rightarrow \mathrm{Var}(X) = \mathbb{E}\,\mathrm{Var}(X\mid Y) + \mathrm{Var}(\mathbb{E}[X\mid Y])$

"within-group variance" + "between-group variance"

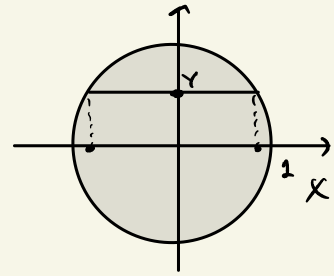
HW: True/false?

- $\mathrm{Var}(X) \geq \mathbb{E}[\mathrm{Var}(X\mid\mathcal{F})]$
- $\mathrm{Var}(X) \geq \mathrm{Var}(X\mid\mathcal{F})$ a.s.

HW: state, prove the Law of Total Covariance

Conditional distributions

- **Intuitive examples**:
  - $X|Y = \begin{cases} 2Y \text{ with prob } 1/2 \\ Y/2 \text{ with prob } 1/2 \end{cases}$ in the Exchange Paradox.
  - $X|Y \sim$ distribution of income (X) in a given country (Y)
  - $Y \sim \mathrm{Unif}[0,1]$; $\quad X|Y \sim N(0, Y^2)$.
  - $(X,Y) \sim \mathrm{Unif}(\text{unit disc}) \Rightarrow X|Y \sim \mathrm{Unif}\left(-\sqrt{1-Y^2}, \sqrt{1-Y^2}\right)$



- **Heuristic def**: the conditional distr. "$X|\mathcal{F}$" is the random prob. meas.

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

- Slight issue: defined a.s — up to a null set.
  The null set may depend on $B \Rightarrow$ not clear why prob. measure.
  However, this issue can be fixed:

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

(i) $\forall \omega \in \Omega$: $\quad B \mapsto \mu(B,\omega)$ is a probability measure on $\mathcal{B}(\mathbb{R})$ ("conditional distr." of X given $\mathcal{F}$)

(ii) $\forall B \in \mathcal{B}(\mathbb{R})$: $\mu(B,\omega) \overset{a.s.}{=\!=} \mathbb{P}\{X \in B \mid \mathcal{F}\}(\omega)$

Proof If Thm holds, $\forall \omega \in \Omega$ the cdf of $\mu(\cdot,\omega)$ must be

$$F(x,\omega) = \mu(-\infty, x] \overset{a.s.}{=} \mathbb{P}\{X \le x \mid \mathcal{F}\}(\omega) = \mathbb{E}\left[\mathbb{1}_{\{X\le x\}} \mid \mathcal{F}\right](\omega).$$

So, let's define the function

$$F(r,\omega) := \mathbb{E}\left[\mathbb{1}_{\{X\le r\}} \mid \mathcal{F}\right](\omega) \qquad \forall r \in \mathbb{Q} \tag{$\circ$}$$

$\forall$ version.

- Excluding null sets, we can find a "nice" set $A \subset \Omega$ with $\mathbb{P}(X\in A)=1$ and such that $\forall \omega \in A$ we have:
  (a) $F(r,\omega) < F(s,\omega)$ $\forall r<s$ in $\mathbb{Q}$ (by monotonicity of cond $\mathbb{E}$)
  (b) $F(r+\frac{1}{n},\omega) \downarrow F(r,\omega)$ as $n\to\infty$, $\forall r \in \mathbb{Q}$ (by cond MCT)
  (c) $F(-n,\omega) \to 0$, $F(n,\omega) \to \infty$ as $n \to \infty$ (by cond MCT)

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

$$F(x,\omega) := \inf\{F(r,\omega) : r > x, r \in \mathbb{Q}\} \quad \forall x \in \mathbb{R} \tag{$\circ\circ$}$$

(Properties (a), (c) show agreement on $\mathbb{Q}$).

And for each $\omega \notin A$, let F be some fixed cdf, e.g.

$$F(x,\omega) := \Phi(x) \quad \forall x \in \mathbb{R}$$

$N(0,1)$

- F is monotonely nondecreasing (by def $\circ\circ$), has the correct limits (0 at $-\infty$, 1 at ∞) (by (c)), and is right-continuous (by (b), def $\circ\circ$ & monotonicity - check!)

$\Rightarrow \forall \omega \in \Omega$: $\quad x \mapsto F(x,\omega)$ is a CDF

$\Rightarrow \forall \omega \in \Omega$: $\exists$ Borel prob measure $B \mapsto \mu(B,\omega)$ with cdf $x \mapsto F(x,\omega)$.

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

(i) the r.v. $\omega \mapsto \mu(B,\omega)$ is $\mathcal{F}$-mble

(ii) $\int_E \mu(B,\omega)\, dP(\omega) = \int_E \mathbb{1}_{\{X(\omega)\in B\}}\, dP(\omega) \qquad \forall\, E \in \mathcal{F}$

(i) $\mathcal{L} := \{ B \in \mathcal{B}(\mathbb{R}) : \omega \mapsto \mu(B,\omega) \text{ is } \mathcal{F}\text{-mble} \}$

is a λ-system $\left(\mu(\cdot,\omega) \text{ is a prob. meas} \Rightarrow \mu\left(\bigsqcup_1^\infty B_i, \omega\right) = \sum_1^\infty \mu(B_i,\omega) \right.$
$\left. \Rightarrow \text{if each function in RHS is } \mathcal{F}\text{-mble} \Rightarrow \text{LHS is } \mathcal{F}\text{-mble} \right)$

that contains the π-system $\mathcal{P} := \{ (-\infty, r],\ r \in \mathbb{Q} \}$ by (o)

π-λ theorem $\Rightarrow$ $\underbrace{\sigma(\mathcal{P})}_{= \mathcal{B}(\mathbb{R})} \subset \mathcal{L}$. ✓

(ii) Fix $\forall\, E \in \mathcal{F}$. Each side of (ii) defines a finite measure of B (by MCT), and identity (ii) holds $\forall B \in \mathcal{P}$ by (o):

$$\int_E \mu((-\infty, r], \omega)\, dP(\omega) = \int_E F(r,\omega)\, dP(\omega) \underset{\substack{\text{by (o) \& def of} \\ \text{condit. exp.}}}{=} \mathbb{E}\left[\mathbb{1}_{\{X \le r\}} \mathbb{1}_E \right].$$

By the uniqueness of measures (or by applying π-λ thm),
(ii) holds $\forall\, B \in \mathcal{B}(\mathbb{R})$. $\square$

HW Extend the thm for r. vectors $X \in \mathbb{R}^n$ [e.g. for $n=2$, work with $\{X \le x, Y \le y\}$ instead of $\{X \le x\}$]

HW (The expectation of the conditional distr = conditional expectation)
Let μ_ω be the conditional distr. of X given $\mathcal{F}$, $h: \mathbb{R} \to \mathbb{R}$ integrable
Show that $\int h(x)\, d\mu_\omega(x) \overset{a.s.}{=} \mathbb{E}[h(X) \mid \mathcal{F}](\omega)$

[redacted]

Conditional densities

- When $\mathcal{F} = \sigma(Y)$, we write $X|Y$ instead of $X|\sigma(Y)$

Prop (Conditional density) If (X,Y) has density $f_{X,Y}(x,y)$, then the conditional distr. of X given Y has density $x \mapsto f_{X|Y}(x|Y)$, where

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

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

• It remains to check that the cond. exp. indeed has this density, i.e.

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

def of cond. exp. $\iff$ $\begin{cases} \int_B f_{X|Y}(x,Y)\,dx \text{ is } \sigma(Y)\text{-mble? Yes! (it's a function of } Y) \\ \mathbb{E}\left[\mathbb{1}_{\{X\in B\}} \mathbb{1}_F\right] \overset{?}{=} \mathbb{E}\left[\left(\int_B f_{X|Y}(x,Y)\,dx\right)\mathbb{1}_F\right] \quad \forall F \in \sigma(Y) \end{cases}$

$F = \{Y \in R\}$ for some $R \in \mathcal{B}(\mathbb{R})$

$$P\{X \in B, Y \in R\} \overset{?}{=} \int_R \left(\int_B f_{X|Y}(x,y)\,dx\right) f_Y(y)\,dy = \int_{R\times B} f(x,y)\,dx\,dy. \quad \text{Yes!}$$

$\square$

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

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

SKIP: "Continuous" version of LTP: $f_X(x) \overset{\text{Lec. 10}}{=} \int_{-\infty}^{\infty} f(x,y)\,dy \overset{\text{Prop}}{=} \int_{-\infty}^{\infty} f_{X|Y}(x|y)\,f_Y(y)\,dy$

$\Rightarrow$ "Continuous" Bayes formula:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \boxed{\frac{f_{Y|X}(y|x)\,f_X(x)}{f_Y(y)}}$$

[Williams 15.7]

Remarks 1. Extensions for r. vectors X, Y – obvious.

2. For discrete X, Y, the same formulas hold for pmfs (easier!)

Application: revisit the filtering problem (p. 152.5):

Signal $X \sim N(0,1)$
noise $W \sim N(0,\sigma^2)$ indep
observation $Y = X + W$.
$X|Y = ?$

Solution: Let's apply Bayes formula.

$$f_X(x) = \text{pdf of } N(0,1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

$$f_{Y|X}(y|x) = \text{pdf of } N(x,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y-x)^2}{2\sigma^2}\right]$$

$$\Rightarrow \log f_{X|Y}(x|y) = \log f_{Y|X}(y|x) + \log f_X(x) + c_1(y)$$

$$= -\frac{(y-x)^2}{2\sigma^2} - \frac{x^2}{2} + c_2(y)$$

$$= -\frac{1+\sigma^2}{2\sigma^2}\left(x - \frac{y}{1+\sigma^2}\right)^2 + c_3(y) \quad \text{(complete the square)}$$

$$\Rightarrow f_{X|Y}(x|y) = \text{pdf of } N\left(\frac{y}{1+\sigma^2}, \frac{\sigma^2}{1+\sigma^2}\right)$$

$\Rightarrow$ the best estimator is $\boxed{\hat{X} = \frac{Y}{1+\sigma^2}}$. Its $\boxed{\text{MSE} = \text{Var}(\hat{X}) = \frac{\sigma^2}{1+\sigma^2}}$

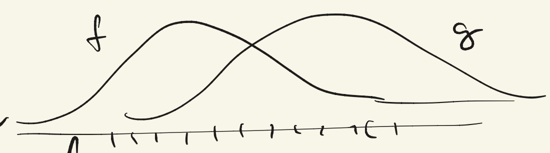
We recovered the result on p. 152.5.

APPLICATION: LIKELIHOOD RATIO TEST

Hypothesis test:

- Given iid observations $X_1, \dots, X_n$, which of the two given distributions have they been sampled from?

- let's say, one has pdf f and the other, g



- Model: $\theta \sim \text{Ber}(1/2)$. Conditionally on θ,

  If $\theta = 0 \Rightarrow$ draw $X = (X_1, \dots, X_n)$ from f

  If $\theta = 1 \Rightarrow$ from g.

- Test: compute the likelihood ratio

$$\Lambda(X) := \frac{P\{\theta = 0 \mid X\}}{P\{\theta = 1 \mid X\}}$$

Output "f" if $\Lambda(X) \geq 1$ and "g" if $\Lambda(X) < 1$

- How to compute $\Lambda(x)$? Bayes:

$$P\{\theta = 0 \mid X = x\} = \frac{P\{X = x \mid \theta = 0\} \cdot P\{\theta = 0\}}{P\{X = x\}}, \qquad x = (x_1, \dots, x_n) \in \mathbb{R}^n$$

and similarly for $P\{\theta = 1 \mid X = x\}$. Divide $\Rightarrow$

$$\frac{P\{\theta = 0 \mid X = x\}}{P\{\theta = 1 \mid X = x\}} = \frac{P\{X = x \mid \theta = 0\}}{P\{X = x \mid \theta = 1\}} = \frac{f(x_1) \cdots f(x_n)}{g(x_1) \cdots g(x_n)}$$

Thus (by "conditioning = fixing" before) $\Rightarrow$

$$\Lambda(X) = \frac{f(X_1) \cdots f(X_n)}{g(X_1) \cdots g(X_n)}.$$

(i.e. "choose f or g, whichever makes the sample $X_1 \dots X_n$ more likely")

Remarks

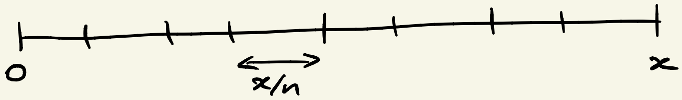
1. For pdf's — the same —
2. LRT is the best test (Neyman-Pearson lemma)
3. Works for > 2 hypotheses: choose the one that maximizes the likelihood of sample (MLE).

SKIP

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

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

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

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

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

$\Downarrow$

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

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

Notation: $X \sim \mathrm{Exp}(\lambda)$. 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.)

SKIP

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

$$P\{X > t+s \mid X > t\} = P\{X > s\} \quad \forall s,t>0$$

$P\{\text{wait} > s \text{ more minutes}\}$ $\quad$ $P\{\text{wait} > s \text{ minutes}\}$

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

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

- **Ex①** Let $X_i \sim Exp(\lambda_i)$, $i=1,2$, be independent.

(a) $P\{X_1 \leq X_2\} = \underset{X_1}{E}[X_1 \leq X_2 \mid X_1] \overset{\text{tail of } X_2}{=} E_{X_1}[e^{-\lambda_2 X_1}] = \int_0^\infty e^{-\lambda_2 x} \cdot \underbrace{\lambda_1 e^{-\lambda_1 x}}_{\text{pdf of } X_1} dx = \boxed{\frac{\lambda_1}{\lambda_1+\lambda_2}}$

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

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

induction $\Rightarrow$ $\boxed{\text{Cor} \quad \min(X_1, \dots, X_n) \sim Exp(\lambda_1 + \dots + \lambda_n)}$ (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 $Exp(\lambda_1)$, $Exp(\lambda_2)$.
Find the expected $\underbrace{\text{time you spend in the office}}_{=: T}$.

$R_i :=$ remaining service time of the customer with clerk i
$\sim Exp(\lambda_i)$ by memoryless property.
$S :=$ your service time. $\quad T = \min(R_1, R_2) + S \quad \Rightarrow$

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

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

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

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

- ~~$EX = \overset{S}{1 \cdot p} + \overset{FS}{2(1-p)p} + \overset{FFS}{3(1-p)^2 p} + \overset{FFFS}{4(1-p)^3 p} + \dots = \frac{1}{p}$ (check!)~~

- Alternative computation:

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

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

($X|F = X+1$ 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

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

SKIP

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

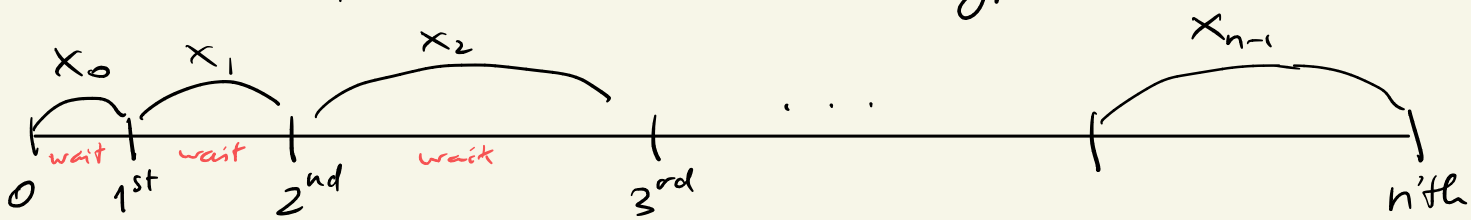
Ex (Coupon collector's problem)

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

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

$n-i$ ← # new types

n ← # all types

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

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

- 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.