Probability & Statistics for EECS **Name (Print):** 柳润康
Spring 2023
Final Exam
June 01, 2023
Time Limit: 180 Minutes Advisor Name ____________

This exam contains 11 pages (including this cover page) and 9 problems. Check to see if any pages are missing. Enter all requested information on the top of this page, and put your initials on the top of every page, in case the pages become separated.

Try to answer as many problems as you can. The following rules apply:

- **Mysterious or unsupported answers will NOT receive full credit**. Correct answers, unsupported by calculations, explanation, or algebraic work will receive no credit. Incorrect answers supported by substantially correct calculations and explanations might still receive partial credit.

Do not write in the table to the right.

Problem	Points	Score
1	5	
2	10	
3	15	
4	20	
5	10	
6	10	
7	10	
8	10	
9	10	
Total:	100	

1. (5 points) Please describe the pros and cons of *Bayesian statistical inference* and *Classical statistical inference*. Then explain why conjugate priors are important for Bayesian statistical inference.

2. (10 points) Let X and Y be two continuous random variables with joint PDF

$$f_{X,Y}(x,y) = \begin{cases} cx^2y & \text{if } 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

   (a) (5 points) Find the value of constant c.

   (b) (5 points) Find the conditional probability $P(Y \le \frac{X}{4} | Y \le \frac{X}{2})$.

3. (15 points) Let X and Y be two integer random variables with joint PMF

$$P_{X,Y}(x,y) = \begin{cases} \frac{1}{6 \cdot 2^{\min(x,y)}} & \text{if } x, y \geq 0, |x-y| \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) (5 points) Find the marginal distributions of X and Y. Are X and Y independent?

(b) (5 points) Find $P(X = Y)$.

(c) (5 points) Find $E[X|Y = 2]$ and $\mathrm{Var}[X|Y = 2]$.

Let X and Y be two integer random variables with joint PMF

$$P_{X,Y}(x,y) = \begin{cases} \frac{1}{6 \cdot 2^{\min(x,y)}}, & \text{if } x, y \geq 0, |x-y| \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find the marginal distributions of X and Y.

(b) Are X and Y independent?

(c) Find $P(X = Y)$.

Solution:

(a) The marginal distributions of X is

$$P_X(X) = \sum_{y=0}^{\infty} P_{X,Y}(X, Y).$$

When $X = 0$, we have

$$P(X = 0) = P(X = 0, Y = 0) + P(X = 0, Y = 1) = \frac{1}{3}.$$

When $X \neq 0$, we have

$$P(X = x) = P(X = x, Y = x - 1) + P(X = x, Y = x) + P(X = x, Y = x + 1) = \frac{1}{6 \cdot 2^{x-2}}.$$

Thus, the marginal distribution of X is

$$P_X(X) = \begin{cases} \frac{1}{3}, & x = 0 \\ \frac{1}{6 \cdot 2^{x-2}}, & x > 0 \\ 0, & \text{otherwise.} \end{cases}$$

According to the symmetric, the marginal distribution of Y is

$$P_Y(Y) = \begin{cases} \frac{1}{3}, & y = 0 \\ \frac{1}{6 \cdot 2^{y-2}}, & y > 0 \\ 0, & \text{otherwise.} \end{cases}$$

(b) Since that

$$P_{X,Y}(0, 0) = \frac{1}{6}, \tag{3}$$

and

$$P(X = 0)P(Y = 0) = \frac{1}{9}, \tag{4}$$

X and Y are not independent.

(c) According to symmetric, we have $P(X = Y) = P(X = Y - 1) = P(X = Y + 1)$ and $P(X = Y) + P(X = Y - 1) + P(X = Y + 1) = 1$. Thus, we have

$$P(X = Y) = \frac{1}{3}.$$

4. (20 points) Let Z_1, Z_2 be two *i.i.d.* random variables satisfying standard normal distributions, *i.e.*, $Z_1, Z_2 \sim \mathcal{N}(0,1)$. Define

$$X = \sigma_X Z_1 + \mu_X;$$
$$Y = \sigma_Y \left(\rho Z_1 + \sqrt{1-\rho^2} Z_2\right) + \mu_Y,$$

where $\sigma_X > 0$, $\sigma_Y > 0$, $-1 < \rho < 1$.

(a) (5 points) Show that X and Y are bivariate normal.

(b) (5 points) Find the correlation coefficient between X and Y, *i.e.*, $\mathrm{Corr}(X,Y)$.

(c) (5 points) Find the joint PDF of X and Y.

(d) (5 points) Find $E[Y|X]$ and $\mathrm{Var}[Y|X]$.

(a) For $a, b \in \mathbb{R}$, we have

$$aX + bY = (a\mathbf{\Sigma}_X + b\mathbf{\Sigma}_Y\rho)Z_1 + b\sqrt{1-\rho^2}\mathbf{\Sigma}_Y Z_2 + a\mu_X + b\mu_Y.$$

Since the linear combination of two Normal distribution follows Normal distribution, X and Y are bivariate normal.

(b) Since $Z_1, Z_2 \sim \mathcal{N}(0,1)$. We have $\rho Z_1 + \sqrt{1-\rho^2} Z_2 \sim \mathcal{N}(0,1)$. So $X \sim \mathcal{N}(\mu_X, \mathbf{\Sigma}_X)$, $Y \sim \mathcal{N}(\mu_Y, \mathbf{\Sigma}_Y)$. Thus, we have

$$\begin{aligned} Cov(X,Y) &= Cov(\mathbf{\Sigma}_X Z_1 + \mu_X, \mathbf{\Sigma}_Y(\rho Z_1 + \sqrt{1-\rho^2}Z_2) + \mu_Y) \\ &= \mathbf{\Sigma}_X\mathbf{\Sigma}_Y Cov(Z_1, \rho Z_1 + \sqrt{1-\rho^2}Z_2) \\ &= \mathbf{\Sigma}_X\mathbf{\Sigma}_Y \left(\rho Var(Z_1) + \sqrt{1-\rho^2} Cov(Z_1, Z_2)\right) \\ &= \mathbf{\Sigma}_X\mathbf{\Sigma}_Y\rho. \end{aligned}$$

Then correlation coefficient between X and y is

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{\mathbf{\Sigma}_X\mathbf{\Sigma}_Y\rho}{\mathbf{\Sigma}_X\mathbf{\Sigma}_Y}.$$

(c) Since Z_1 and Z_2 are i.i.d., we have

$$f_{Z_1,Z_2}(z_1,z_2) = f_{Z_1}(z_1)f_{Z_2}(z_2) = \frac{1}{2\pi}e^{-\frac{z_1^2+z_2^2}{2}}.$$

Since $X = \mathbf{\Sigma}_X Z_1 + \mu_X$, $Y = \mathbf{\Sigma}_Y(\rho Z_1 + \sqrt{1-\rho^2}Z_2) + \mu_Y$, we have

$$Z_1 = \frac{X - \mu_X}{\mathbf{\Sigma}_X}$$

and

$$Z_2 = \frac{Y-\mu_Y}{\sqrt{1-\rho^2}\mathbf{\Sigma}_Y} - \rho\frac{X-\mu_X}{\sqrt{1-\rho^2}\mathbf{\Sigma}_X}.$$

Thus,

$$\begin{aligned} f_{X,Y}(x,y) &= \left|\frac{\partial(Z_1,Z_2)}{\partial(X,Y)}\right| f_{Z_1,Z_2}(z_1,z_2) \\ &= \frac{1}{\begin{vmatrix} \frac{\partial z_1}{\partial x} & \frac{\partial z_1}{\partial y} \\ \frac{\partial z_2}{\partial x} & r\frac{\partial z_2}{\partial x} \end{vmatrix}} f_{Z_1,Z_2}(z_1,z_2) \\ &= \frac{1}{\begin{vmatrix} \frac{1}{\mathbf{\Sigma}_X} & 0 \\ \frac{-\rho}{\sqrt{1-\rho^2}\mathbf{\Sigma}_X} & \frac{1}{\sqrt{1-\rho^2}\mathbf{\Sigma}_Y} \end{vmatrix}} f_{Z_1,Z_2}(z_1,z_2) \\ &= \frac{1}{\mathbf{\Sigma}_X\mathbf{\Sigma}_Y\sqrt{1-\rho^2}} f_{Z_1,Z_2}(z_1,z_2) \\ &= \frac{1}{\mathbf{\Sigma}_X\mathbf{\Sigma}_Y\sqrt{1-\rho^2}} f_{Z_1,Z_2}\left(\frac{x-\mu_X}{\mathbf{\Sigma}_X}, \frac{y-\mu_Y}{\sqrt{1-\rho^2}\mathbf{\Sigma}_Y} - \rho\frac{x-\mu_X}{\sqrt{1-\rho^2}\mathbf{\Sigma}_X}\right) \\ &= \frac{1}{2\pi\mathbf{\Sigma}_X\mathbf{\Sigma}_Y\sqrt{1-\rho^2}} e^{-\frac{(\frac{x-\mu_X}{\mathbf{\Sigma}_X})^2 + (\frac{y-\mu_Y}{\sqrt{1-\rho^2}\mathbf{\Sigma}_Y} - \rho\frac{x-\mu_X}{\sqrt{1-\rho^2}\mathbf{\Sigma}_X})^2}{2}} \\ &= \frac{1}{2\pi\mathbf{\Sigma}_X\mathbf{\Sigma}_Y\sqrt{1-\rho^2}} e^{-\frac{(\frac{x-\mu_X}{\mathbf{\Sigma}_X})^2 - \frac{2\rho(x-\mathbf{\Sigma}_X)(Y-\mathbf{\Sigma}_Y)}{\mathbf{\Sigma}_X\mathbf{\Sigma}_Y} + (\frac{y-\mu_Y}{\mathbf{\Sigma}_Y})^2}{2(1-\rho^2)}}. \end{aligned}$$

5. (10 points) Let the random variable $X \sim \mathcal{N}(\mu, \tau^2)$. Given $X = x$, random variables $Y_1, Y_2, \ldots, Y_n$ are *i.i.d.* and have the same conditional distribution, *i.e.*, $Y_i | X = x \sim \mathcal{N}(x, \sigma^2)$. Define the sample mean $\bar{Y}$ as follows:

$$\bar{Y} = \frac{Y_1 + \cdots + Y_n}{n}.$$

   (a) (4 points) Find the posterior PDF of X given $\bar{Y}$.

   (b) (3 points) Find the MAP (Maximum a Posterior Probability) estimates of X given $\bar{Y}$.

   (c) (3 points) Find the MMSE estimates of X given $\bar{Y}$. (We know that the MMSE of X given Y is given by $g(Y) = E[X|Y]$).

6. (10 points) Let $X_1 \sim \text{Expo}(\lambda_1)$, $X_2 \sim \text{Expo}(\lambda_2)$ and $X_3 \sim \text{Expo}(\lambda_3)$ be independent.
   (a) (5 points) Find $E(X_1 + X_2 + X_3 | X_1 > 1, X_2 > 2, X_3 > 3)$ in terms of $\lambda_1, \lambda_2, \lambda_3$.
   (b) (5 points) Find $P(X_1 = \min(X_1, X_2, X_3))$.

$E(X_i | X_i > t) = t + E(X_i)$

(a) $E(X_1 | X_1 > 1) = 1 + \frac{1}{\lambda_1}$

$E(X_2 | X_2 > 2) = 2 + \frac{1}{\lambda_2}$

$E(X_3 | X_3 > 3) = 3 + \frac{1}{\lambda_3}$

$= (1 + \frac{1}{\lambda_1}) + ( )$

(b) $P(X_i = \min(X_1, X_2 \cdots X_n)) = \frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j}$

$P(X_1 = \min(X_1, X_2, X_3)) = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3}$

7. (10 points) Let $X \sim \text{Expo}(\lambda)$, $Y \sim \text{Expo}(\lambda)$; X and Y are independent.
   (a) (5 points) Find $E(X|X+Y)$.
   (b) (5 points) Find $E(X^2|X+Y)$.

(a) $Z = X+Y \sim \text{Gamma}(2,\lambda)$ $E(Z) = \frac{2}{\lambda}$

$E(X|X+Y) + E(Y|X+Y) = E(X+Y|X+Y) = E(Z|Z) = \frac{2}{\lambda}$

$E(X|X+Y) = \frac{1}{\lambda}$

Other method: $f_{X|X+Y}(x|x+y) \sim \text{Unif}(0, x+y)$

$E(X|X+Y) = \frac{X+Y}{2}$

$E(E(X|X+Y)) = E(\frac{X+Y}{2}) = E(\frac{X}{2}) + E(\frac{Y}{2}) = \frac{1}{2\lambda} + \frac{1}{2\lambda} = \frac{1}{\lambda}$

$E(X|Z) = \int_0^z x \frac{1}{z} dx = \frac{1}{z} \cdot \frac{z^2}{2} = \frac{z}{2}$

$E(E(X|Z)) = E(\frac{Z}{2}) = \frac{1}{2} \times \frac{2}{\lambda} = \frac{1}{\lambda}$

(b) Similarly, $f_{X|X+Y}(x|x+y) \sim \text{Unif}(0, x+y)$

Let $Z = X+Y$ $f_{X|Z}(x|z) \sim \text{Unif}(0,z)$

$E(X^2|Z) = \int_0^z x^2 \cdot \frac{1}{z} dx = \frac{z^2}{3}$

$E(E(X^2|Z)) = E(\frac{Z^2}{3}) = \frac{1}{3}(\frac{2}{\lambda^2} + \frac{2}{\lambda^2}) = \frac{4}{3\lambda^2}$

$f_X(x) = \lambda e^{-\lambda x}$ $f_Y(y) = \lambda e^{-\lambda y}$ $(x, y > 0)$

$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y) = \lambda^2 e^{-\lambda(x+y)}$

Define $U = X+Y$ $V = \frac{X}{X+Y}$

$\Rightarrow \begin{cases} X = UV \\ Y = U(1-V) \end{cases}$ $\Rightarrow J = |U|$

joint PDF $f_{U,V}(u,v) = f_{X,Y}(uv, u(1-v)) \cdot |J|$

$= \lambda^2 e^{-\lambda u} \cdot u$ $\left(\begin{array}{l} u>0 \\ 0 \le v \le 1 \end{array} \right)$

marginal PDF $f_U(u) = \int_0^1 \lambda^2 e^{-\lambda u} \cdot u \, dv$

$= \lambda^2 e^{-\lambda u} \cdot u$

conditional PDF $f_{X|U}(x|u) = \frac{f_{U,V}(u, \frac{x}{u})}{}$

X given X+Y can be viewed as uniform distribution on $[0, X+Y]$ $E(X|X+Y) = \frac{X+Y}{2}$

$E[g(x)] = \int_{-\infty}^{\infty} g(x) f(x) dx$

$g(x) = x^2$

$E(X^2|X+Y) = \int_0^{x+y} \frac{1}{x+y} \cdot x^2 dx$

$= \frac{(x+y)^2}{3}$

$\text{Cov}(X, X+Y) = \text{Cov}(X,X) + \text{Cov}(X,Y) = \text{Var}(X) + 0 = \frac{1}{\lambda^2}$

8. (10 points) Instead of predicting a single value for the parameter, we give an interval that is likely to contain the parameter: A $1-\delta$ confidence interval for a parameter p is an interval $[\hat{p}-\epsilon, \hat{p}+\epsilon]$ such that $P(p \in [\hat{p}-\epsilon, \hat{p}+\epsilon]) \geq 1-\delta$. Now we toss a coin with probability p landing heads and probability $1-p$ landing tails. The parameter p is unknown and we need to estimate its value from experiment results. We toss such coin N times. Let $X_i = 1$ if the ith result is head, otherwise 0. We estimate p by using $\hat{p} = \frac{X_1+\cdots+X_N}{N}$. Find the $1-\delta$ confidence interval for p, then discuss the impacts of δ and N. **Hint**: You can use the following Hoeffding bound: Let the random variables $X_1, X_2, \ldots, X_n$ be independent with $E(X_i) = \mu$, $a \leq X_i \leq b$ for each $i = 1, \ldots, n$, where a, b are constants. Then for any $\epsilon \geq 0$,

$E(X_i)=\mu$ $a \le X_i \le b$

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \epsilon\right) \leq 2\exp\left(-\frac{2n\epsilon^2}{(b-a)^2}\right).$$

$E(\hat{p}) = p$. $Var(\hat{p}) = \frac{p(1-p)}{N}$

Chebyshev inequality
$P(|X-\mu| \geq a) \leq \frac{\sigma^2}{a^2}$

Chernoff's inequality. for $a, t>0$
$P(X \geq a) \leq \frac{E(e^{tX})}{e^{ta}}$

Hoeffding Bound.

$\Leftrightarrow$ $P(|\hat{p}-p| \geq \varepsilon) \leq \delta$

Chebyshev inequality: $P(|\hat{p}-p| \geq \varepsilon) \leq \frac{Var(\hat{p})}{a^2} = \frac{p(1-p)}{N\varepsilon^2}$

$\delta = \frac{p(1-p)}{N\varepsilon^2} \Rightarrow \varepsilon = \sqrt{\frac{p(1-p)}{N\delta}}$

impact: 1. fix δ. reducing error ε requires Larger N
2. fix ε. narrowing confidence interval $\to$ larger N

Hoeffding Bound: $P(|\hat{p}-p| \geq \varepsilon) \leq 2e^{-2N\varepsilon^2}$

$\Rightarrow \delta = 2e^{-2N\varepsilon^2}$. $\varepsilon = \sqrt{\frac{\ln(\frac{2}{\delta})}{2N}}$

9. (10 points) Show the following inequalities.

(a) (5 points) Let $X \sim \text{Pois}(\lambda)$. If there exists a constant $a > \lambda$, then

$$P(X \geq a) \leq \frac{e^{-\lambda}(e\lambda)^a}{a^a}.$$

(b) (5 points) Let X be a random variable with finite variance σ^2. Then for any constant $a > 0$,

$$P(|X - \mathbb{E}[X]| \geq a) \leq \frac{2\sigma^2}{\sigma^2 + a^2}.$$

Chernoff's Inequality:

(a) $P(X \geq a) \leq \frac{E(e^{tX})}{e^{ta}}$ for any r.v. X and any $t>0$

if $X \sim Pois(\lambda)$, then $M_X(t)$ is given by $E[e^{tX}] = e^{\lambda(e^t-1)}$

so we have $P(X \geq a) \leq \frac{e^{\lambda(e^t-1)}}{e^{ta}}$

Let $f(t) = \frac{e^{\lambda(e^t-1)}}{e^{ta}} = \exp(\lambda(e^t-1) - ta)$

set $f'(t) = (\lambda e^t - a)\exp(\cdots) = 0 \Rightarrow t = \ln\frac{a}{\lambda}$

since $f''(t) > 0$, then we can conclude that

$f(t)$ get the minimum when $t = \ln\frac{a}{\lambda}$

whose value is $f(\ln\frac{a}{\lambda}) = \frac{e^{\lambda(\frac{a}{\lambda}-1)}}{(e^{\ln\frac{a}{\lambda}})^a} = \frac{e^{a-\lambda}}{(\frac{a}{\lambda})^a} = \frac{e^{-\lambda}(e\lambda)^a}{a^a}$

i.e. $f(t)_{min} = \frac{e^{-\lambda}(e\lambda)^a}{a^a}$

since we have proved that $P(X \geq a) \leq f(t)$

Thus we can conclude that $P(X \geq a) \leq \frac{e^{-\lambda}(e\lambda)^a}{a^a}$

(b) Define $Y = X - E[X]$, then we have $E[Y] = 0$, $Var(Y) = \sigma^2$. for any $u \geq 0$, we have

$P(X - E[X] \geq a) = P(Y \geq a) = P(Y + u \geq a + u) \leq P((Y+u)^2 \geq (a+u)^2)$

since $P((Y+u)^2 \geq (a+u)^2) \leq \frac{E[(Y+u)^2]}{(a+u)^2} = \frac{Var(Y+u) + [E(Y+u)]^2}{(a+u)^2} = \frac{\sigma^2 + u^2}{(a+u)^2}$ (Markov's Inequality)

Let $f(u) = \frac{\sigma^2 + u^2}{(a+u)^2}$ $f'(u) = \frac{2au}{(a+u)^3} - \frac{2\sigma^2}{(a+u)^3}$ set $f'(u) = 0 \Rightarrow u_0 = \frac{\sigma^2}{a^2}$ (easy to find that $f''(u_0) > 0$)

so $f(u)_{min} = f(u_0) = \frac{\sigma^2 + \frac{\sigma^4}{a^4}}{(a + \frac{\sigma^2}{a})^2} = \frac{\sigma^2 + \frac{\sigma^4}{a^4}}{a^2 + \frac{\sigma^4}{a^2} + 2\sigma^2}$

Thus to conclude that $P(Y \geq a) \leq \frac{2\sigma^2}{\sigma^2 + a^2}$

	Y discrete	**Y continuous**
X discrete	$P(Y=y\|X=x)=\frac{P(X=x\|Y=y)P(Y=y)}{P(X=x)}$	$f_Y(y\|X=x)=\frac{P(X=x\|Y=y)f_Y(y)}{P(X=x)}$
X continuous	$P(Y=y\|X=x)=\frac{f_X(x\|Y=y)P(Y=y)}{f_X(x)}$	$f_{Y\|X}(y\|x)=\frac{f_{X\|Y}(x\|y)f_Y(y)}{f_X(x)}$

	Y discrete	**Y continuous**
X discrete	$P(X=x)=\sum_y P(X=x\|Y=y)P(Y=y)$	$P(X=x)=\int_{-\infty}^{\infty} P(X=x\|Y=y)f_Y(y)dy$
X continuous	$f_X(x)=\sum_y f_X(x\|Y=y)P(Y=y)$	$f_X(x)=\int_{-\infty}^{\infty} f_{X\|Y}(x\|y)f_Y(y)dy$

Figure 1: Bayes' Rule & LOTP.

Table of distributions

Name	**Param.**	**PMF or PDF**	**Mean**	**Variance**
Bernoulli	p	$P(X=1)=p, P(X=0)=q$	p	pq
Binomial	n,p	$\binom{n}{k}p^k q^{n-k}$, for $k\in\{0,1,\dots,n\}$	np	npq
FS	p	pq^{k-1}, for $k\in\{1,2,\dots\}$	$1/p$	q/p^2
Geom	p	pq^k, for $k\in\{0,1,2,\dots\}$	q/p	q/p^2
NBinom	r,p	$\binom{r+n-1}{r-1}p^r q^n, n\in\{0,1,2,\dots\}$	rq/p	rq/p^2
HGeom	w,b,n	$\frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}$, for $k\in\{0,1,\dots,n\}$	$\mu=\frac{nw}{w+b}$	$(\frac{w+b-n}{w+b-1})n\frac{\mu}{n}(1-\frac{\mu}{n})$
Poisson	λ	$\frac{e^{-\lambda}\lambda^k}{k!}$, for $k\in\{0,1,2,\dots\}$	λ	λ
Uniform	$a<b$	$\frac{1}{b-a}$, for $x\in(a,b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Normal	μ,σ^2	$\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/(2\sigma^2)}$	μ	σ^2
Log-Normal	μ,σ^2	$\frac{1}{x\sigma\sqrt{2\pi}}e^{-(\log x-\mu)^2/(2\sigma^2)}, x>0$	$\theta=e^{\mu+\sigma^2/2}$	$\theta^2(e^{\sigma^2}-1)$
Expo	λ	$\lambda e^{-\lambda x}$, for $x>0$	$1/\lambda$	$1/\lambda^2$
Gamma	a,λ	$\Gamma(a)^{-1}(\lambda x)^a e^{-\lambda x}x^{-1}$, for $x>0$	a/λ	a/λ^2
Beta	a,b	$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}$, for $0<x<1$	$\mu=\frac{a}{a+b}$	$\frac{\mu(1-\mu)}{a+b+1}$
Chi-Square	n	$\frac{1}{2^{n/2}\Gamma(n/2)}x^{n/2-1}e^{-x/2}$, for $x>0$	n	$2n$
Student-t	n	$\frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)}(1+x^2/n)^{-(n+1)/2}$	0 if $n>1$	$\frac{n}{n-2}$ if $n>2$

	Y **discrete**	Y **continuous**
X **discrete**	$P(Y=y\|X=x)=\frac{P(X=x\|Y=y)P(Y=y)}{P(X=x)}$	$f_Y(y\|X=x)=\frac{P(X=x\|Y=y)f_Y(y)}{P(X=x)}$
X **continuous**	$P(Y=y\|X=x)=\frac{f_X(x\|Y=y)P(Y=y)}{f_X(x)}$	$f_{Y\|X}(y\|x)=\frac{f_{X\|Y}(x\|y)f_Y(y)}{f_X(x)}$

	Y **discrete**	Y **continuous**
X **discrete**	$P(X=x)=\sum_y P(X=x\|Y=y)P(Y=y)$	$P(X=x)=\int_{-\infty}^{\infty} P(X=x\|Y=y)f_Y(y)dy$
X **continuous**	$f_X(x)=\sum_y f_X(x\|Y=y)P(Y=y)$	$f_X(x)=\int_{-\infty}^{\infty} f_{X\|Y}(x\|y)f_Y(y)dy$

Figure 1: Bayes' Rule & LOTP.

Table of distributions

Name	**Param.**	**PMF or PDF**	**Mean**	**Variance**
Bernoulli	p	$P(X=1)=p, P(X=0)=q$	p	pq
Binomial	n,p	$\binom{n}{k}p^k q^{n-k}$, for $k\in\{0,1,\dots,n\}$	np	npq
FS	p	pq^{k-1}, for $k\in\{1,2,\dots\}$	$1/p$	q/p^2
Geom	p	pq^k, for $k\in\{0,1,2,\dots\}$	q/p	q/p^2
NBinom	r,p	$\binom{r+n-1}{r-1}p^r q^n, n\in\{0,1,2,\dots\}$	rq/p	rq/p^2
HGeom	w,b,n	$\frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}$, for $k\in\{0,1,\dots,n\}$	$\mu=\frac{nw}{w+b}$	$(\frac{w+b-n}{w+b-1})n\frac{\mu}{n}(1-\frac{\mu}{n})$
Poisson	λ	$\frac{e^{-\lambda}\lambda^k}{k!}$, for $k\in\{0,1,2,\dots\}$	λ	λ
Uniform	$a<b$	$\frac{1}{b-a}$, for $x\in(a,b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Normal	μ,σ^2	$\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/(2\sigma^2)}$	μ	σ^2
Log-Normal	μ,σ^2	$\frac{1}{x\sigma\sqrt{2\pi}}e^{-(\log x-\mu)^2/(2\sigma^2)}, x>0$	$\theta=e^{\mu+\sigma^2/2}$	$\theta^2(e^{\sigma^2}-1)$
Expo	λ	$\lambda e^{-\lambda x}$, for $x>0$	$1/\lambda$	$1/\lambda^2$
Gamma	a,λ	$\Gamma(a)^{-1}(\lambda x)^a e^{-\lambda x}x^{-1}$, for $x>0$	a/λ	a/λ^2
Beta	a,b	$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}$, for $0<x<1$	$\mu=\frac{a}{a+b}$	$\frac{\mu(1-\mu)}{a+b+1}$
Chi-Square	n	$\frac{1}{2^{n/2}\Gamma(n/2)}x^{n/2-1}e^{-x/2}$, for $x>0$	n	$2n$
Student-t	n	$\frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)}(1+x^2/n)^{-(n+1)/2}$	0 if $n>1$	$\frac{n}{n-2}$ if $n>2$