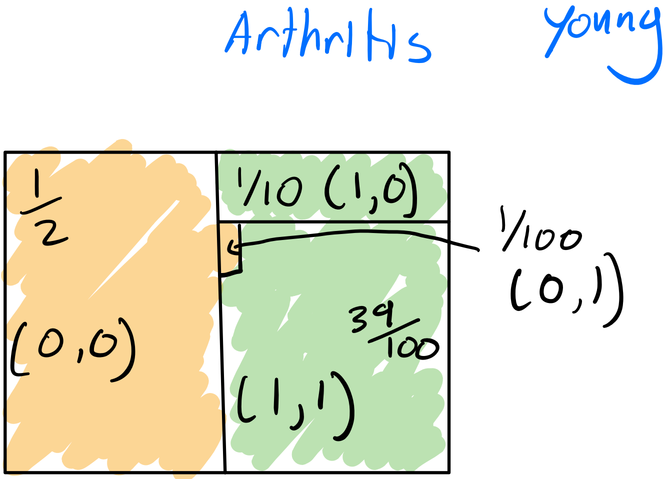
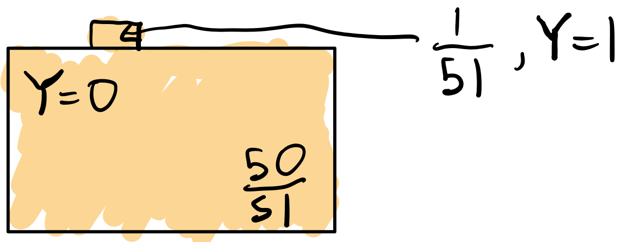
Important Announcements

- assignment 1 is due this Friday (Jan 24)
- review session by TA on Monday



$\downarrow$ condition on being young



$$= P_{Y|X}(1|0)$$

product rule from above
$p(x,y) = p(x|y)\,p(y)$

$$= \frac{P_{X,Y}(0,1)}{P_X(0)}$$

$$= \frac{p(0,1)}{p(0,1)+p(0,0)}$$

$$= \frac{1/100}{1/100 + 1/2}$$

$$= \frac{1/100}{51/100} = \frac{1}{51}$$

<u>Ex</u>: Probability of being young given I have arthritis

$$\mathbb{P}(X=0 \mid Y=1) = P_{X|Y}(0|1)$$

$$\overset{\text{Bayes' Rule}}{=} \frac{P_{Y|X}(1|0)\,P_X(0)}{P_Y(1)}$$

$$= \frac{\frac{1}{51}\;\frac{51}{100}}{40/100} = \frac{1}{40}$$

<u>Independence</u>: Changing the value of one r.v. doesn't affect the probability of another r.v.

r.v. X, Y are independent if: $p(x,y) = p(x)p(y)$

Since $p(x,y) = p(x|y)p(y) = p(x)p(y|x) = p(x)p(y)$

independence implies: $p(x|y) = p(x)$, $p(y|x) = p(y)$

More generally:

$X_1, X_2, \dots, X_d$ are independent if: $p(x_1, \dots, x_d) = p(x_1) \cdots p(x_d)$

Similarly for distributions:

r.v. X, Y are independent if: $\mathbb{P}(X \in \mathcal{E}_X, Y \in \mathcal{E}_Y) = \mathbb{P}(X \in \mathcal{E}_X)\mathbb{P}(Y \in \mathcal{E}_Y)$

Ex: X, Y are <u>not</u> independent for Arthritis ex

$$p(0,1) = \frac{1}{100} \neq p_X(0)p_Y(1) = \frac{51}{100}\,\frac{40}{100} = 0.204$$

Ex: $X_1, X_2 \in \{0,1\}$ are flips of two different fair coins

$$p(x_1, x_2) = \frac{1}{4} \quad \text{for all } x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2$$

$$p_{X_1}(x_1)p_{X_2}(x_2) = \frac{1}{2}\,\frac{1}{2} = \frac{1}{4}$$

X_1 \ X_2	H	T
H	1/4	1/4
T	1/4	1/4

What happens when $Z = (X, Y)$ with Y discrete and X continuous?

$p : \mathcal{X} \times \mathcal{Y} \Rightarrow ?$ pmf or pdf? Ans: neither

Instead we will write $p(x,y)$ in terms of a marginal pdf for X and a conditional pmf for $Y|X$

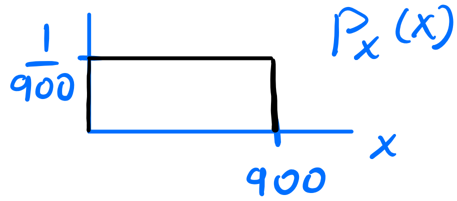
$$p(x,y) = p_X(x)\, p_{Y|X}(y|x)$$ product rule

where $p_{Y|X=x} : \mathcal{Y} \Rightarrow [0,1]$ is a pmf

$p_X(x) : \mathcal{X} \Rightarrow [0,\infty)$ is a pdf

Ex: $X \in \mathcal{X} = [0, 900]$, $Y \in \mathcal{Y} \in \{0, 1\}$ Barolo

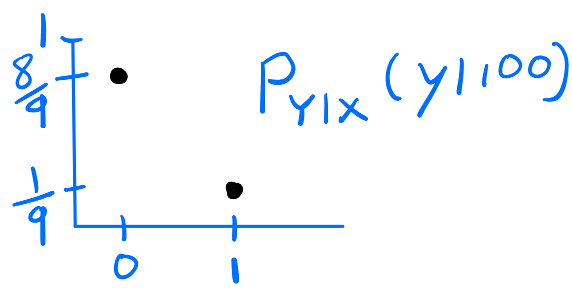
pdf: $p_X = \text{Uniform}(0, 900)$
$= \frac{1}{900}$



$$P_{Y|X=x} = \text{Bernoilli}\left(\frac{x}{900}\right)$$

pmf: $$P_{Y|X}(y|x) = \begin{cases} \frac{x}{900} & \text{if } y = 1 \\ 1 - \frac{x}{900} & \text{if } y = 0 \end{cases}$$

Defn of Bernoilli$\left(\frac{x}{900}\right)$



$$\mathbb{P}(X \in [0, 50], Y = 1) = \int_0^{50} \left(\sum_{y \in \{1\}} p(y, x) \right) dx$$

$$= \int_0^{50} \left(\sum_{y \in \{1\}} P_{Y|X}(y|x)\, p_X(x) \right) dx$$

$$= \int_0^{50} p_{Y|X}(1|x)\, p_X(x)\, dx$$

$$= \int_0^{50} \frac{x}{900} \frac{1}{900}\, dx$$

$$= \frac{1}{810000} \left. \frac{x^2}{2} \right|_0^{50}$$

$$= \frac{1}{810000} \frac{2500}{2}$$

$$= \frac{1}{648}$$

0.154%

Representing Random Features, Labels, and Datasets

Random variables:

$$D = (Z_1, Z_2, \dots, Z_n) \in \mathcal{Z}_1 \times \dots \times \mathcal{Z}_n = \mathcal{Z}^n \quad \text{since } \mathcal{Z} = \mathcal{Z}_1 = \dots = \mathcal{Z}_n$$

$$Z_i = (\vec{X}_i, Y_i) \in \mathcal{X} \times \mathcal{Y} = \mathcal{Z}$$ each Z_i is a feature-label pair

$$\vec{X}_i = (X_{i,1}, \dots, X_{i,d})^T \in \mathbb{R}^d = \mathcal{X}$$ $\vec{X}_i$ is a feature vector

Distributions:

$\mathbb{P}_D$: distribution for D, $\mathbb{P}_{Z_i}$: marginal distribution for Z_i

assumptions:

1. $(\vec{X}_i, Y_i) = Z_i$ are indendent for all $i \in \{1, \dots, n\}$

2. $\mathbb{P}_{Z_1} = \mathbb{P}_{Z_2} = \dots = \mathbb{P}_{Z_n} = \mathbb{P}_Z$ all Z_i have the same distribution

"$(\vec{X}_i, Y_i)$ are independent and identically distributed (i.i.d)"

$$\mathbb{P}_D(Z_1 \in E_1, \dots, Z_n \in E_n) = \mathbb{P}_{Z_1}(Z_1 \in E_1) \cdots \mathbb{P}_{Z}(Z_n \in E_n)$$

$$= \mathbb{P}_{Z}(Z_1 \in E_1) \cdots \mathbb{P}_{Z}(Z_n \in E_n)$$

Equivalently:

$$D = ((\vec{X}_1, Y_1), \dots, (\vec{X}_n, Y_n)) \in (\mathcal{X} \times \mathcal{Y})^n$$

where $(\vec{X}_i, Y_i) \sim \mathbb{P}_{\vec{X}, Y}$ are idependent for all $i \in \{1, \dots, n\}$

"sampled/distributed according to"

D contains n independent samples of $(\vec{X}_i, Y_i)$ "feature-label" pairs all coming from the same distribution $\mathbb{P}_{\vec{X},Y}$

Functions of Random Variables

A function of a r.v. is a r.v.

Ex: (X is a fair six-sided dice)

$X \in \{1,2,3,4,5,6\} = \mathcal{X}$ with $p(x) = \frac{1}{6}$

$f(X) = X^2 \in \underbrace{\{1^2, 2^2, 3^2, 4^2, 5^2, 6^2\}}_{\text{outcome space for } f(X)} = \mathcal{Y}$ is a r.v.

Notice $f: \mathcal{X} \to \mathcal{Y}$

Sometimes we give the r.v. a new symbol

$Y = f(X) = X^2$

(pmf) $P_Y(y) = P_{f(X)}(y) = \frac{1}{6}$ where $y \in \{1, 2^2, 3^2, 4^2, 5^2, 6^2\}$

In this case $p_Y(x^2) = p(x)$ where $x \in \mathcal{X}$

ex: $p_Y(9) = p_Y(3^2) = p(3) = \frac{1}{6}$

Ex: (X is the payout from a slot machine)

$X \in [-10, 10]$ with $p(x) = \frac{1}{20}$, $\mathbb{P} =$ Uniform $(-10, 10)$

$Y = f(X) = X^2 \in [0, 100] = \mathcal{Y}$

$p_Y(y) = \frac{1}{20\sqrt{Y}}$ much more complicated

In general p_Y is complicated and we will not need to know how to calculate it

The Predictor and Learner are functions of r.v.

Ex: (Predictor)

$\vec{X} = (X_1, X_2)^T \in \mathbb{R}^2 = \mathcal{X}$ with $\mathbb{P}_{\vec{X}}$

predictor: $f: \mathcal{X} \to \mathcal{Y}$ where $\mathcal{Y} = \mathbb{R}$

$f(\vec{X}) = 3 + 6X_1 + 2.5X_2$ is a r.v. with values in $\mathcal{Y}$

and has some distribution $\mathbb{P}_{f(\vec{X})}$

Ex: (Learner)

$D = ((\vec{X}_1, Y_1), \ldots, (\vec{X}_n, Y_n)) \in (\mathcal{X} \times \mathcal{Y})^n$ with $\mathbb{P}_D$

Learner: $\mathcal{A} : (\mathcal{X} \times \mathcal{Y})^n \Rightarrow \{f \mid f : \mathcal{X} \Rightarrow \mathcal{Y}\} = \mathcal{F}$

$\mathcal{A}(D) = f$ is a r.v. with values in $\mathcal{F}$

and has some distribution $\mathbb{P}_{\mathcal{A}(D)}$

example

if $D = ((7, 6), (12, 2.5))$ where $n = 2$, $\mathcal{X} = \mathbb{R}$, $\mathcal{Y} = \mathbb{R}$

then f_D can be $f(x) = 2.5 + 6x$

This means we can talk about things like:

- What is the probability the Predictor $f(\vec{x})$ outputs some value y

- What is the probability the Learner $\mathcal{A}(D)$ outputs some predictor f

Expectation and Variance

Expected Value of a r.v.: average value of the r.v. if you sample from its distribution infinitely many times.

The r.v. <u>must</u> take values in $\mathbb{R}$.

It is <u>not</u> always the value we expect to see most frequently (that is the mode)

$X \in \mathcal{X}$ is a r.v. with pmf or pdf p

$$\boxed{\mathbb{E}[X] \overset{\text{def}}{=} \begin{cases} \sum_{x \in \mathcal{X}} x\, p(x) & \text{if } X \text{ is discrete} \\ \int_{\mathcal{X}} x\, p(x)\, dx & \text{if } X \text{ is continuous} \end{cases}}$$

Ex: (fair six-sided dice)

$X \in \{1,2,3,4,5,6\} = \mathcal{X}$ and $\mathbb{P} = \text{Uniform}(n=6)$

thus $p(x) = \frac{1}{6}$

$$\mathbb{E}[X] = \sum_{x \in \mathcal{X}} x\, p(x) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6}$$

$$= 3.5$$

This is <u>not</u> an number you can roll on a dice!

Ex: (Unfair coin)

$X \in \{0,1\}$ and $\mathbb{P} = \text{Bernoulli}(\alpha)$

thus $p(1) = \alpha,\ p(0) = 1-\alpha$

$$\mathbb{E}[X] = \sum_{x \in \mathcal{X}} x\, p(x) = 0 \cdot (1-\alpha) + 1 \cdot \alpha$$
$$= \alpha$$

This is <u>not</u> a result of a coin flip (unless $\alpha = 1$ or $\alpha = 0$)

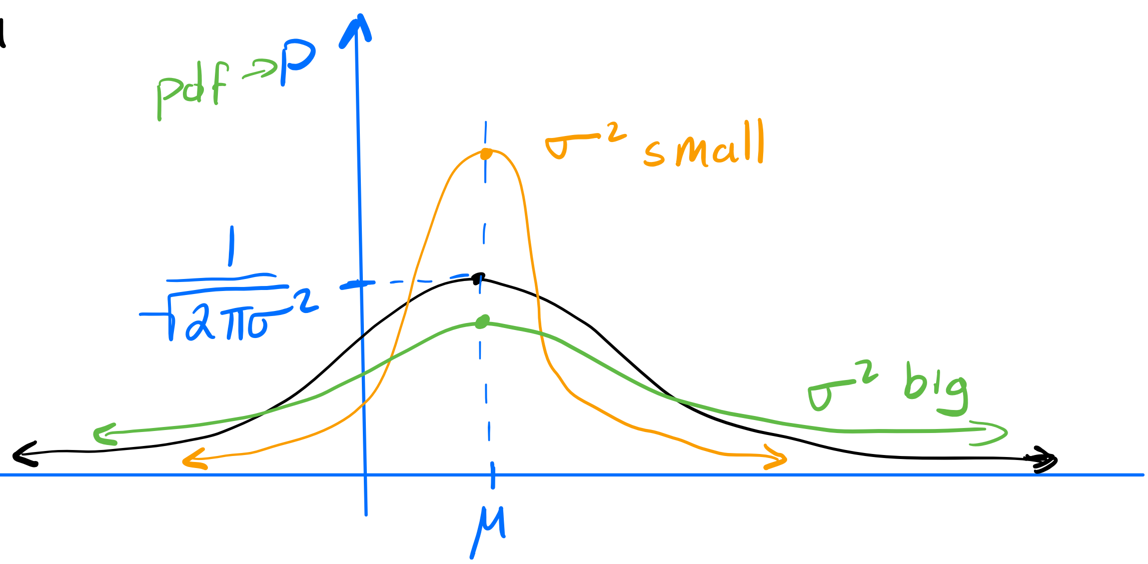
Ex: (Normal distribution)

$X \in \mathbb{R} = \mathcal{X}$ and $\mathbb{P} = \mathcal{N}(\mu, \sigma^2)$

thus $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$

$$\mathbb{E}[X] = \int_{\mathcal{X}} x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx$$

You don't need to know the steps $= \mu$



Expected value of functions of r.v.:

$X \in \mathcal{X}$ is a r.v. with pmf or pdf p

The function $f : \mathcal{X} \to \mathcal{Y}$ must have $\mathcal{Y} = \mathbb{R}$

$$\boxed{\mathbb{E}[f(X)] \stackrel{\text{def}}{=} \begin{cases} \sum_{x \in \mathcal{X}} f(x)p(x) & \text{if } X \text{ is discrete} \\ \int_{\mathcal{X}} f(x)p(x)\,dx & \text{if } X \text{ is continuous} \end{cases}}$$

Ex: (X is the payout from a slot machine)

$X \in [-10, 10]$ with $p(x) = \frac{1}{20}$, $\mathbb{P} = \text{Uniform}(-10, 10)$

$Y = f(X) = X^2 \in [0, 100] = \mathcal{Y}$

$p_Y(y) = \frac{1}{20\sqrt{Y}}$ much more complicated

$$\mathbb{E}[f(X)] = \int_{\mathcal{X}} f(x)p(x)\,dx = \int_{-10}^{10} x^2 \frac{1}{20}\,dx$$

$$= \frac{x^3}{3} \cdot \frac{1}{20} \Big|_{-10}^{10}$$

$$= \left(\frac{1000}{3} - \left(\frac{-1000}{3}\right)\right) \cdot \frac{1}{20}$$

$$= \frac{2000}{60} = 33.333$$

It turns out

$$\mathbb{E}[f(X)] = \mathbb{E}[Y] = \int_y y \, p_Y(y) \, dy$$

$$= 33.333$$

Usually we don't know $p_Y = p_{f(X)}$

So we work with p

Variance of a r.v.: How much the r.v. varies from its expected value on average

$X \in \mathcal{X}$ is a r.v. with pmf or pdf p

$$\boxed{\mathrm{Var}[X] \stackrel{\text{def}}{=} \mathbb{E}\Big[\underbrace{\big(X - \mathbb{E}[X]\big)^2}\Big] = \mathbb{E}[X^2] - \big(\mathbb{E}[X]\big)^2}$$

this is just a function of the r.v. X

Ex: (Unfair coin)

$X \in \{0, 1\}$ and $\mathbb{P} = \text{Bernoulli}(\alpha)$

thus $p(1) = \alpha, p(0) = 1 - \alpha$

$$\mathbb{E}[X] = \sum_{x \in \mathcal{X}} x\, p(x) = 0 \cdot (1-\alpha) + 1 \cdot \alpha$$
$$= \alpha$$

$$\mathrm{Var}[X] = \mathbb{E}\left[\left(X - \mathbb{E}[X]\right)^2\right]$$
$$= \sum_{x \in \mathcal{X}} (x - \mathbb{E}[X])^2 p(x)$$
$$= (0-\alpha)^2 \cdot (1-\alpha) + (1-\alpha)^2 \cdot \alpha$$
$$= \alpha^2 - \alpha^3 + \alpha - 2\alpha^2 + \alpha^3$$
$$= \alpha - \alpha^2$$
$$= \alpha(1-\alpha)$$

or $\mathrm{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

$$= \sum_{x \in \mathcal{X}} x^2 p(x) - \alpha^2$$
$$= 0^2 \cdot (1-\alpha) + 1^2 \cdot \alpha - \alpha^2$$
$$= \alpha(1-\alpha)$$

Ex: (Normal distribution)

$X \in \mathbb{R} = \mathcal{X}$ and $\mathbb{P} = \mathcal{N}(\mu, \sigma^2)$

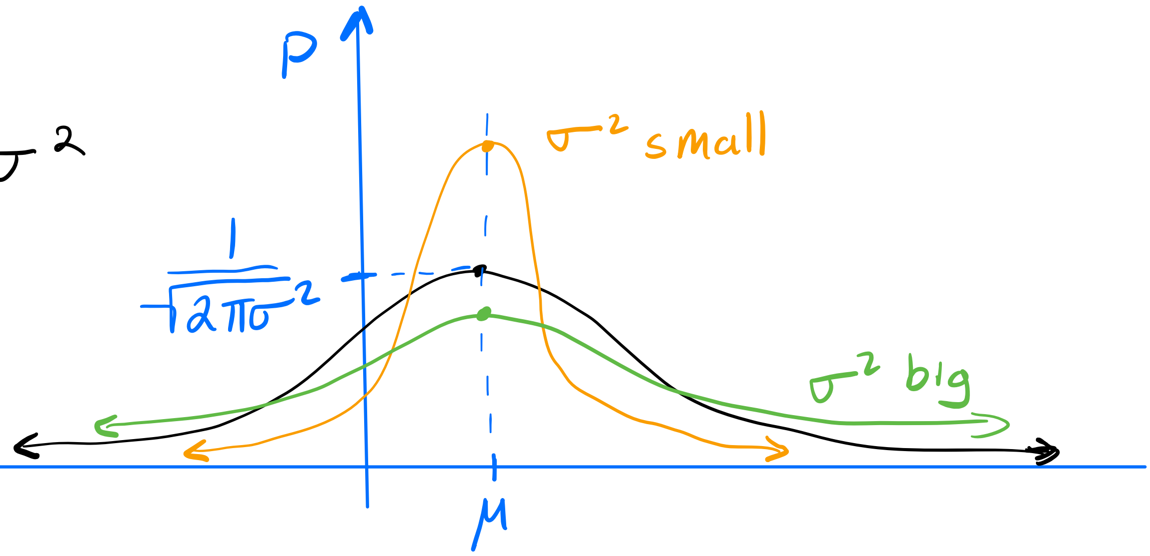
thus $p(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\dfrac{1}{2\sigma^2}(x-\mu)^2\right)$

$$\mathbb{E}[X] = \int_{\mathcal{X}} x \, \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx$$

You don't need to know the steps

$$= \mu$$

$$\mathrm{Var}[X] = \sigma^2$$



Multivariate Expected Value:

$Z = (X, Y) \in \mathcal{X} \times \mathcal{Y} = \mathcal{Z}$ is a r.v.

$f : \mathcal{X} \times \mathcal{Y} \Rightarrow \mathbb{R}$

$$\mathbb{E}[f(X,Y)] = \begin{cases} \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} f(x,y) p(x,y) & \text{if } X, Y \text{ are discrete} \\ \int_{\mathcal{X}} \int_{\mathcal{Y}} f(x,y) p(x,y) \, dy \, dx & \text{if } X, Y \text{ are continuous} \\ \int_{\mathcal{X}} \left(\sum_{y \in \mathcal{Y}} f(x,y) p(y|x) \right) p(x) \, dx & \text{if } Y \text{ is discrete and } X \text{ is continuous} \\ \sum_{x \in \mathcal{X}} \left(\int_{\mathcal{Y}} f(x,y) p(y|x) \, dy \right) p(x) & \text{if } Y \text{ is continuous and } X \text{ is discrete} \end{cases}$$

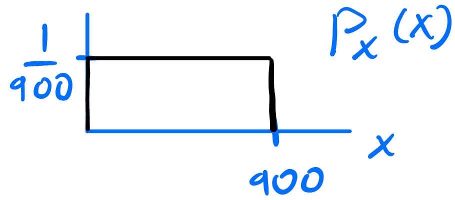
you can always use:

$$p(x,y) = p(y|x)\,p(x) = p(x|y)\,p(x)$$

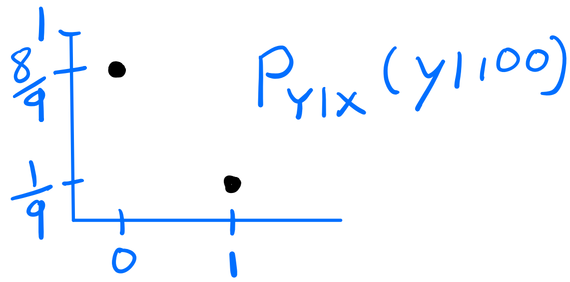
$\underline{\underline{\text{Ex}}}$: $X \in \mathcal{X} = [0, 900]$, $Y \in \mathcal{Y} \in \{0, 1\}$ Barolo

pdf: $p_X = \text{Uniform}(0, 900)$

$\qquad = \frac{1}{900}$



$P_{Y|X=x} = \text{Bernoilli}\left(\frac{x}{900}\right)$



pmf: $P_{Y|X}(y|x) = \begin{cases} \frac{x}{900} & \text{if } y = 1 \\ 1 - \frac{x}{900} & \text{if } y = 0 \end{cases}$

Defn of Bernoilli$\left(\frac{x}{900}\right)$

$$f(X,Y) = \left(\frac{X}{900} - Y\right)^2$$

$$\mathbb{E}[f(X,Y)] = \int_{\mathcal{X}} \left(\sum_{y \in \mathcal{Y}} f(x,y)\,p(y|x)\right) p(x)\,dx$$

$$= \int_0^{900} \left(\sum_{y \in \{0,1\}} \left(\frac{x}{900} - y\right)^2 p(y|x)\right) p(x)\,dx$$

$$= \int_0^{900} \left(\left(\frac{x}{900} - 0\right)^2 \left(1 - \frac{x}{900}\right) + \left(\frac{x}{900} - 1\right)^2 \left(\frac{x}{900}\right)\right) \frac{1}{900}\,dx$$

$$= \frac{1}{900} \int_0^{900} \frac{x}{900} \left(1 - \frac{x}{900}\right) dx$$

$$= \frac{1}{900} \left(\frac{x^2}{1800} \Big|_0^{900} - \frac{x^3}{3 \cdot 900^2} \Big|_0^{900} \right)$$

$$= \frac{1}{6}$$

Conditional Expected Value:

$(X, Y) \in \mathcal{X} \times \mathcal{Y}$ is a r.v.

$p = p_{Y|X}$ is a conditional pmf or pdf

$f : \mathcal{Y} \to \mathbb{R}$

$$\boxed{\mathbb{E}[f(Y) \mid X = x] = \begin{cases} \sum_{y \in \mathcal{Y}} f(y)\, p(y|x) & \text{if } Y \text{ is discrete} \\ \int_{\mathcal{Y}} f(y)\, p(y|x) & \text{if } Y \text{ is continuous} \end{cases}}$$

Useful Properties

Let X, Y be r.v. and $c \in \mathbb{R}$ be a constant

1. $\mathbb{E}[cX] = c\,\mathbb{E}[X]$
2. $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
3. $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]]$
4. $Var[c] = 0$
5. $Var[cX] = c^2\,Var[X]$

will be on the formula sheet for midterms

If X and Y are independent:

6. $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$
7. $Var[X+Y] = Var[X] + Var[Y]$