Lecture 02
August 23, 2023

CE/MATH 362/322
Statistics and Inferencing

Bernoulli Distribution

If X is a random variable with a Bernoulli distribution, then:

$$\Pr(X = 1) = p = 1 - \Pr(X = 0) = 1 - q.$$

The probability mass function f of this distribution, over possible outcomes k, is

$$f(k;p) = \begin{cases} p & \text{if } k = 1, \\ q = 1 - p & \text{if } k = 0. \end{cases}$$

$$\boxed{\begin{aligned} &\mathbb{E}[X] = p \\ &\mathbb{E}[X^2] = p \\ &\operatorname{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = p(1-p) = pq \end{aligned}}$$

Binomial Distribution

If $X \sim \text{Bin}(n, p)$, that is, X is a binomially distributed random variable, n being the total number of experiments and p the probability of each experiment yielding a successful result.

In other words, X is the sum of n identical Bernoulli random variables, each with expected value p. In other words, if $X_1, \ldots, X_n$ are identical (and independent) Bernoulli random variables with parameter p, then $X = X_1 + \cdots + X_n$ is binomially distributed random variable.

$$\boxed{\begin{aligned} \mathrm{E}[X] &= np \\ \mathrm{E}\left[X^2\right] &= np(1-p) + n^2p^2 \\ \mathrm{Var}(X) &= np(1-p) \end{aligned}}$$

Owing to Independence

$$\mathrm{E}[X_i X_j] = \mathrm{E}[X_i]\mathrm{E}[X_j] = (np)(np) = n^2p^2$$

$$
\begin{aligned}
\mathbb{E}\left[X_{\text{bin}}^2\right] &= \mathbb{E}\left[(X_1 + \ldots + X_n)^2\right] \\
&= \mathbb{E}\left[X_1^2 + \ldots + X_n^2 + 2(X_1X_2 + X_1X_3 + \ldots + X_{n-1}X_n)\right] \\
&= \sum_{i=1}^{n} \mathbb{E}\left[X_i^2\right] + 2\sum_{i=1}^{n} \sum_{j=i+1}^{n} \mathbb{E}[X_iX_j] \\
&= n\mathbb{E}\left[X_1^2\right] + 2\binom{n}{2}\mathbb{E}[X_1X_j] \quad (j \neq 1) \\
&= n\mathbb{E}\left[X_1^2\right] + n(n-1)\mathbb{E}[X_1]\mathbb{E}[X_j] \quad (j \neq 1) \\
&= np + n(n-1)p^2 \\
\mathbf{Var}(X_{\text{bin}}) &= \mathbb{E}\left[X_{\text{bin}}^2\right] - (\mathbb{E}[X_{\text{bin}}])^2 = np + n(n-1)p^2 - n^2p^2 = np(1-p)
\end{aligned}
$$

Let X be a Scaled version of Binomially Distributed RV

$$X = \frac{X_{\text{bin}}}{n} = \frac{X_1 + X_2 + \ldots + X_n}{n}, \qquad \text{where } X_{\text{bin}} \sim \text{Bin}(n, p)$$

$$\begin{aligned} \text{E}[X_{\text{bin}}] &= np \\ \text{Var}(X_{\text{bin}}) &= np(1-p) \end{aligned}$$

$$\mathbb{E}[X] = \mathbb{E}\left[\frac{X_{\text{bin}}}{n}\right] = \frac{1}{n}\mathbb{E}[X_{\text{bin}}] = \frac{np}{n} = p$$

$$\text{Var}(X) = \text{Var}\left(\frac{X_{\text{bin}}}{n}\right) = \frac{1}{n^2}\text{Var}(X_{\text{bin}}) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

Polling Survey: Estimating the Percentage of Votes Casted

Let's try to understand how polling survey can be used to determine the popular support of a candidate in some province (say, Punjab).

Key quantities:

$N = 5000,000 =$ population of Punjab (assumed)

$$p = \frac{\text{\# people who support Nawaz Sharif}}{N}$$

$$1 - p = \frac{\text{\# people who support Imran Khan}}{N}$$

We know N but we don't know p.

Question #1: What is p ?
Question #2: Is $p > 0.5$?
Question #3: Are you sure?
Question #4: If a poll is conducted, and n people are voted, where $n < N$;
~~ ~~ ~~ ~~ how good would that estimate of p be?

Suppose we poll a simple random sample of $n = 1000$ people from the population of Punjab. This means:

- Person 1 is chosen at random (equally likely) from all N people in Iowa. Then person 2 is chosen at random from the remaining $N - 1$ people. Then person 3 is chosen at random from the remaining $N - 2$ people, etc.

- Or equivalently, all $\binom{N}{n} = \frac{N!}{n!(N-n)!}$ possible sets of n people are equally likely to be chosen.

Then we can estimate p by

$$\hat{p} = \frac{\#\text{ sampled people who support Nawaz Sharif}}{n}$$

Say 540 out of the 1000 people we surveyed support Nawaz, so $\hat{p} = 0.54$.

Does this mean $p = 0.54$? Does this mean $p > 0.5$?

No! Let's call our data $X_1, \ldots, X_n$:

$$X_i = \begin{cases} 1 & \text{if person } i \text{ supports Nawaz} \\ 0 & \text{if person } i \text{ supports Imran} \end{cases}$$

Then $\hat{p} = \frac{X_1 + X_2 + \ldots + X_n}{n}$

The data $X_1, \ldots, X_n$ are random, because we took a random sample.
Therefore $\hat{p}$ is also random.

$\hat{p}$ is a random variable-it has a probability distribution.
We can ask:
What is $\mathbb{E}[\hat{p}]$?
What is $\mathrm{Var}[\hat{p}]$?
What is the distribution of $\hat{p}$?

Each of the N people of Punjab is equally likely to be the i^{th} person that we sampled. So each $X_i \sim \text{Bernoulli}(p)$, and $\mathbb{E}[X_i] = p$.

$$\mathbb{E}[\hat{p}] = \mathbb{E}\left[\frac{X_1 + \ldots + X_n}{n}\right] = \frac{1}{n}(\mathbb{E}[X_1] + \ldots + \mathbb{E}[X_n]) = p$$

Interpretation: The "average value" of $\hat{p}$ is p.
We say that $\hat{p}$ is **unbiased**.

$$
\begin{aligned}
\mathbb{E}\left[\hat{p}^{\,2}\right] &= \mathbb{E}\left[\left(\frac{X_1 + \ldots + X_n}{n}\right)^2\right] \\
&= \frac{1}{n^2}\mathbb{E}\left[X_1^2 + \ldots + X_n^2 + 2(X_1X_2 + X_1X_3 + \ldots + X_{n-1}X_n)\right] \\
&= \frac{1}{n^2}\left(n\mathbb{E}\left[X_1^2\right] + 2\binom{n}{2}\mathbb{E}[X_1X_j]\right) \quad (j \neq 1) \\
&= \frac{1}{n}\mathbb{E}\left[X_1^2\right] + \frac{n-1}{n}\mathbb{E}[X_1X_j] \quad (j \neq 1)
\end{aligned}
$$

$$
\mathbb{E}\left[{X_i}^2\right] = p, \quad 1 \leqslant i \leqslant n
$$

But how to find $\mathbb{E}[X_iX_j]$? $(i \neq j)$

Q: Are X_1 and X_2 independent?
A: No.

$$\mathbb{E}[X_1 X_2] = \mathbb{P}[X_1 = 1, X_2 = 1] = \mathbb{P}[X_1 = 1]\,\mathbb{P}[X_2 = 1 \mid X_1 = 1]$$

We have:

$$\mathbb{P}[X_1 = 1] = p, \quad \mathbb{P}[X_2 = 1 \mid X_1 = 1] = \frac{Np - 1}{N - 1}$$

$$\mathbb{E}\left[\hat{p}^{\,2}\right] = \frac{1}{n}\mathbb{E}\left[X_1^2\right] + \frac{n-1}{n} \cdot \frac{Np - 1}{N - 1} \cdot p$$

$$
\begin{aligned}
\mathrm{Var}[\hat{p}] &= \mathbb{E}[\hat{p}^2] - (\mathbb{E}[\hat{p}])^2 \\
&= \frac{1}{n}p + \frac{n-1}{n}p\left(\frac{Np-1}{N-1}\right) - p^2 \\
&= \left(\frac{1}{n} - \frac{n-1}{n}\frac{1}{N-1}\right)p + \left(\frac{n-1}{n}\frac{N}{N-1} - 1\right)p^2 \\
&= \frac{N-n}{n(N-1)}p + \frac{n-N}{n(N-1)}p^2 \\
&= \frac{p(1-p)}{n}\frac{N-n}{N-1} = \frac{p(1-p)}{n}\left(1 - \frac{n-1}{N-1}\right)
\end{aligned}
$$

$$\mathrm{Var}[\hat{p}] = \frac{p(1-p)}{n}\left(1 - \frac{n-1}{N-1}\right)$$

What if $N \to \infty$?

$$\mathrm{Var}[\hat{p}] \approx \frac{p(1-p)}{n} \Longrightarrow \hat{p} \longrightarrow p \quad \text{as} \quad n \longrightarrow \infty$$

What if $n = N$?

$$\mathrm{Var}[\hat{p}] = 0 \Longrightarrow \hat{p} = p$$

When N is much bigger than n, this is approximately $\frac{p(1-p)}{n}$, which would be the variance if we sampled n people in Punjab with replacement. (In that case, $\hat{p}$ would be a Binomial(n, p) random variable divided by n). The factor $1 - \frac{n-1}{N-1}$ is the **correction for sampling without replacement.** (More will be discussed later when we cover Chapter 7 of Rice on Survey Sampling.)

For $N = 5,000,000$, $n = 1000$, and $p \approx 0.54$, the standard deviation of $\hat{p}$ is $\sqrt{\mathrm{Var}[\hat{p}]} \approx 0.01575$. So, the estimate of p will lie somewhere in the range 0.54 ± 0.01575, i.e., from 0.52425 to 0.55575 (with a high probability).

If you ask few of your close friends, where $n = 10$, then the standard deviation of $\hat{p}$ is $\sqrt{\mathrm{Var}[\hat{p}]} \approx 0.1576$. So, the estimate of p will lie somewhere in the range 0.54 ± 0.1576, i.e., from 0.3824 to 0.6976 with a high probability.

So, what did you learn today from this example?

Review of Joint, Conditional and Marginal PDF

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

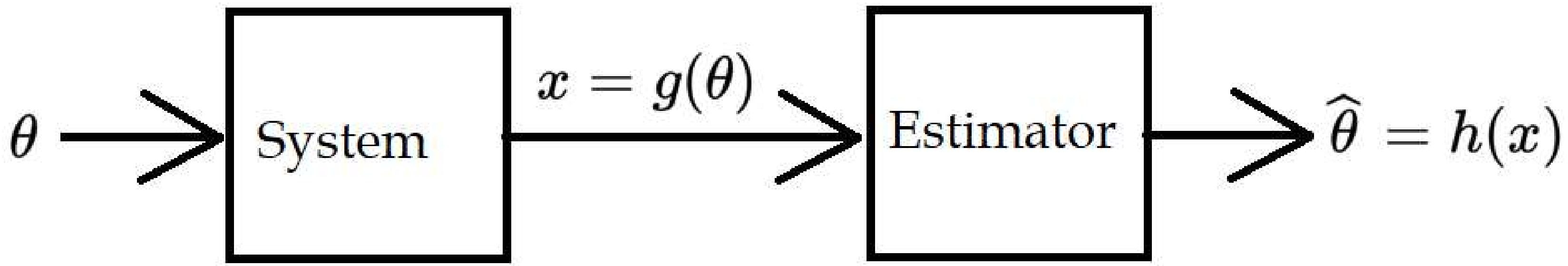
$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy$$

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

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

Bayes Estimator
Minimum mean square error estimation



The most common risk function used for Bayesian estimation is the mean square error (MSE), also called squared error risk. The MSE is defined by

$$\mathrm{MSE} = \mathbb{E}\left[(\widehat{\theta}(x) - \theta)^2\right],$$

where the expectation is taken over the joint distribution of θ and x.

Bayes Estimator
Minimum mean square error estimation

The most common risk function used for Bayesian estimation is the mean square error (MSE), also called squared error risk. The MSE is defined by

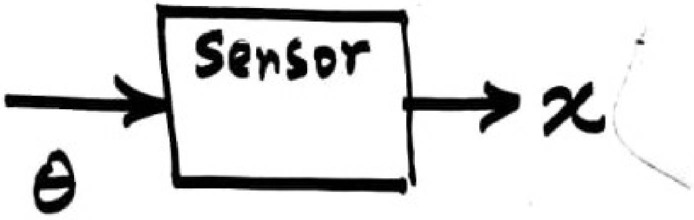
$$\mathrm{MSE} = \mathbb{E}\left[(\widehat{\theta}(x) - \theta)^2\right],$$

where the expectation is taken over the joint distribution of θ and x.

Using the MSE as risk, the Bayes estimate of the unknown (true) parameter θ, denoted as $\widehat{\theta}(x)$ is simply the mean of the posterior distribution, the conditional pdf of θ given x, where $x = g(\theta)$ is a set of observations which depends on θ in some manner:

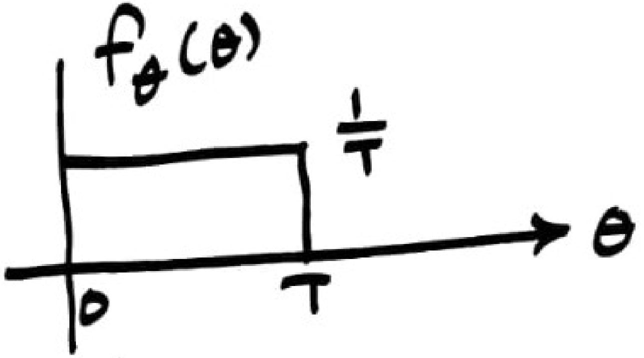
$$\widehat{\theta} = \mathbb{E}[\theta \mid x] = h(x) = \int \theta \, f_{\theta|x}(\theta \mid x) \, d\theta.$$

This is known as the **minimum mean square error (MMSE) estimator**.



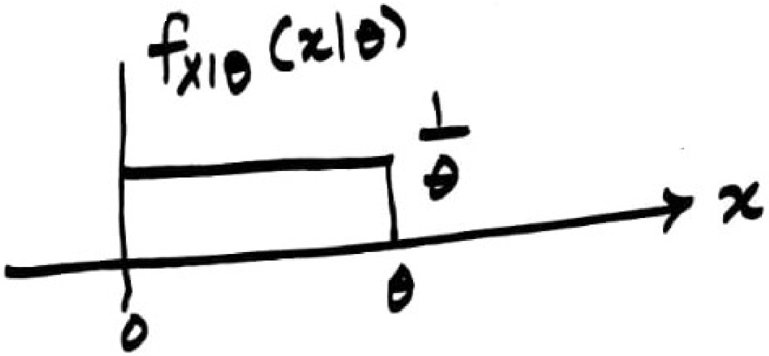
θ is uniformly distributed, $\theta \sim \mathcal{U}[0, T]$

Given θ, the observation x is also uniformly distributed $x|\theta \sim \mathcal{U}[0, \theta]$



$$f_\theta(\theta) = \begin{cases} \frac{1}{T} & 0 \leq \theta \leq T \\ 0 & \text{otherwise} \end{cases}$$

The task is to obtain a Bayes estimator of θ using the observation x



$$f_{X|\theta}(x|\theta) = \begin{cases} \frac{1}{\theta}, \text{ for } 0 \leq x \leq \theta \\ \text{zero, otherwise} \end{cases}$$

We may obtain joint density

$$f_{X,\theta}(x,\theta) = f_{X|\theta}(x|\theta) f_{\theta}(\theta) = \begin{cases} \frac{1}{T\theta} \text{ for } 0 \leq x \leq \theta \leq T \\ \text{zero, otherwise} \end{cases}$$

$$f_X(x) = \int f_{X,\theta}(x,\theta)\, d\theta = \frac{1}{T}\int_x^T \frac{1}{\theta}\, d\theta = \frac{\ln(T) - \ln(x)}{T}$$

for $0 \leq x \leq T$.

In the next class, we will obtain the Bayes estimator of θ.