Recap:

① Cramer-Rao lower bound.

② Multivariate Cramer-Rao lower bound.

Methods of finding estimators

(i) Method of moments

(ii) Method of maximum likelihood estimation

(iii) Bayes and minimax estimators

Method of moments

Sometimes we do not know how to construct estimators for a bunch of parameters and we need to depend on intuitions. In that case, creating a sample analogue of the population parameters might work as an estimator.

In general we create the method of moment estimator as following.

Let $m_j = \frac{1}{n} \sum_{i=1}^{n} x_i^j$,

also assume $\mu_j' = E[X^j]$

Generally μ_j''s are functions of unknown parameters $\theta_1, \ldots, \theta_k$. We find the method of moment estimators (MOM) of $\theta_1, \ldots, \theta_k$

by solving k equations

$$m_j = \mu_j'(\theta_1, \dots, \theta_k) \quad , \; j = 1, \dots, k.$$

<u>Example</u>: $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

We will use, $\quad m_1 = E[X]$

$$m_2 = E[X^2]$$

$$\Rightarrow \quad \frac{1}{n}\sum_{i=1}^{n} X_i = E[X] \quad \text{- - - - - } (1)$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = E[X^2] \quad \text{- - - - - } (2)$$

$$\Rightarrow \quad \frac{1}{n}\sum_{i=1}^{n} X_i = \mu$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = \mu^2 + \sigma^2 \quad \Rightarrow \quad \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i ,$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^2$$

In this case, MOM estimators are functions of complete, sufficient statistics. However, it is not the case in many distributions.

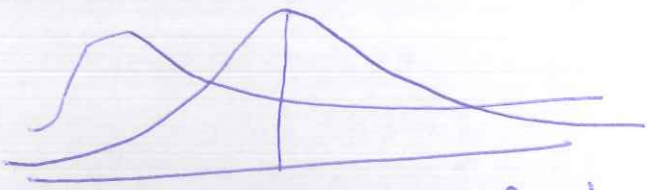
e.g. $\quad X_1, \dots, X_n \overset{iid}{\sim} U(\mu - 2, \mu + 2)$

$$\frac{1}{n}\sum_{i=1}^{n} X_i = E[X] = \mu$$

$\Rightarrow$ MOM estimator of μ is $\bar{X}$.

Here minimal sufficient statistic is a function of $X_{(1)}, X_{(n)}$ .. Hence MOM estimator is not

a function of minimal sufficient statistic.



Another problem of the MOM estimator is that it might not provide estimators in the support of the parameter space.

<u>Example</u>: $X_1, \ldots, X_n \overset{iid}{\sim} Bin(k, p)$, k and p are both unknown.

$$\frac{1}{n}\sum_{i=1}^{n} X_i = E[X] \quad \Rightarrow \quad \frac{1}{n}\sum_{i=1}^{n} X_i = kp$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = E[X^2] \qquad \frac{1}{n}\sum_{i=1}^{n} X_i^2 = kp(1-p) + k^2p^2$$

$$\Rightarrow p = \frac{\bar{X}}{k} \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n} X_i^2 = kp - kp^2 + k^2p^2$$

$$= \bar{X} + kp^2(k-1)$$

$$\Rightarrow \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X} = k\left(\frac{\bar{X}}{k}\right)^2(k-1)$$

$$\Rightarrow \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X} = \frac{\bar{X}^2}{k}(k-1)$$

$$\Rightarrow k\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}\right] = \bar{X}^2 k - \bar{X}^2$$

$$\Rightarrow k = \frac{\bar{X}^2}{\bar{X}^2 - \frac{1}{n}\sum_{i=1}^{n} X_i^2 + \bar{X}}$$

the R.H.S. may not be an integer while L.H.S. is an integer. Hence MOM estimator is not even a feasible estimator here.

Example: $X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, \theta^2)$, $\theta > 0$

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \theta$$

and $\frac{1}{n}\sum_{i=1}^{n} X_i^2 = E[X^2] = \theta^2 + \theta^2 = 2\theta^2$

$\Rightarrow \theta = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\theta = \sqrt{\frac{1}{2n}\sum_{i=1}^{n} X_i^2}$

Maximum Likelihood estimator

Suppose $f_{\theta_1, \ldots, \theta_k}(x)$ be the p.d.f. of X. Then the likelihood of $\underline{X} = (X_1, \ldots, X_n)$ is given by

$$L(\theta_1, \ldots, \theta_k) = \prod_{i=1}^{n} f_{\theta_1, \ldots, \theta_k}(x_i)$$

Maximum Likelihood estimator (MLE) is the value of $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)$ that maximizes the likelihood $L(\theta_1, \ldots, \theta_k)$. MLE is found by solving $\frac{\partial L(\underline{\theta})}{\partial \theta_i} = 0$, $i = 1, \ldots, k$.

sometimes you use $\frac{\partial \log L(\theta)}{\partial \theta_i} = 0$, $i = 1, \ldots, k$

to find the MLE.

Note that $\frac{\partial \log L(\theta)}{\partial \theta_i} = 0 \Rightarrow \frac{1}{L(\theta)} \frac{\partial L(\theta)}{\partial \theta_i} = 0$

$\Rightarrow \frac{\partial L(\theta)}{\partial \theta_i} = 0$. Thus they are equivalent.

$$\frac{\partial \log L(\theta)}{\partial \theta_i} = 0 \Rightarrow \frac{\partial \left[\sum_{j=1}^{n} \log f_{\theta_1, \ldots, \theta_k}(x_j)\right]}{\partial \theta_i} = 0$$

$$\Rightarrow \sum_{j=1}^{n} \frac{\partial \log f_{\theta_1, \ldots, \theta_k}(x_j)}{\partial \theta_i} = 0 \quad , i = 1, \ldots, k.$$

Note that $\frac{\partial \log f_{\theta_1,\dots,\theta_k}(x_j)}{\partial \theta_i}$ for $i=1,\dots,k$ are components of the score vector.

In the one dimensional case where $k=1$, the equations boil down to

$$\sum_{j=1}^{n} \frac{\partial \log f_\theta(x_j)}{\partial \theta} = 0 \Rightarrow \sum_{j=1}^{n} u_\theta(x_j) = 0$$

Let $\hat{\theta}_1, \dots, \hat{\theta}_k$ be the solutions of the system of equations to find the MLE. The MLE should also satisfy that $\left.\frac{\partial^2 L(\underline{\theta})}{\partial \underline{\theta}^2}\right|_{\theta=\hat{\theta}} < 0$

MLE sometimes lies in the boundary of the parameter space.

<u>Example</u>: $X_1, \dots, X_n \overset{iid}{\sim} N(\theta, 1)$, $\theta > 0$

Find the MLE of θ.

When $\theta > 0$ restriction is not there, MLE of θ is given by $\bar{X}$. However, when we have this additional restriction we can't use $\bar{X}$ as the MLE.

If $\bar{X} > 0$, then obviously $\bar{X}$ is going to be the MLE.

If $\bar{x} \leq 0$ then $\bar{x}$ is not the MLE.

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(x_i-\theta)^2}{2}\right\}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\sum_{i=1}^{n}\frac{(x_i-\theta)^2}{2}\right\}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{n}(x_i-\bar{x}+\bar{x}-\theta)^2\right]\right\}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{n}(x_i-\bar{x})^2+(\bar{x}-\theta)^2+2(x_i-\bar{x})(\bar{x}-\theta)\right]\right\}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^2\right\} \exp\left\{-\frac{n}{2}(\bar{x}-\theta)^2\right\}$$

Note that $\bar{x} \leq 0$ so, let $\bar{x} = -z$, $z \geq 0$

$$L(\theta) \propto \exp\left\{-\frac{n}{2}(-z-\theta)^2\right\}$$

$$= \exp\left\{-\frac{n}{2}(z+\theta)^2\right\}$$

Note that the range of θ is $\theta > 0$.

As θ increases the function $L(\theta)$ monotonically decreases.

The maximum is achieved when $\theta = 0$ which is outside the range of θ.

Thus no MLE exists in the case when $\bar{x} \leq 0$.

However if we begin with the distribution

$X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, 1), \quad \theta \geq 0$

then when $\bar{x} \geq 0$ $\theta = \bar{x}$ maximizes the likelihood.

When $\bar{x} < 0$, the likelihood is maximized at $\theta = 0$.

MLE of θ in this problem is $\max\{0, \bar{x}\}$

<u>Remark:</u> MLE does not need to be a function of minimal sufficient statistic.

$X_1, \ldots, X_n \overset{iid}{\sim} U(\theta - 1, \theta + 1)$

$$L(\theta) = \frac{1}{2^n} I\left[X_{(n)} - 1 < \theta < X_{(1)} + 1\right]$$

Any value of $\theta \in (X_{(n)} - 1, X_{(1)} + 1)$ maximizes the likelihood. Hence MLE is not unique.

Also,

$$\hat{\theta} = \left(\frac{\bar{x}}{\bar{x}+1}\right)(X_{(n)} - 1) + \frac{1}{(\bar{x}+1)}(X_{(1)} + 1)$$

$\hat{\theta} \in (X_{(n)} - 1, X_{(1)} + 1)$ being a weighted average of $X_{(n)} - 1$ and $X_{(1)} + 1$.

Thus $\hat{\theta}$ is an MLE, but it is not a function of a minimal sufficient statistic.

Invariance property of MLE

If $\hat{\theta}$ is an MLE of θ, $\tau(\hat{\theta})$ is an MLE of $\tau(\theta)$, for any function $\tau(\theta)$.

Recap: ① Method of moments estimator

② Maximum Likelihood estimator.

Invariance property of MLE:

If $\hat{\theta}$ is the MLE of θ then $\tau(\hat{\theta})$ is the MLE of $\tau(\theta)$.

Asymptotic distribution of MLE:

In smooth regular problems MLE $\hat{\theta}$ has the property that $\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} N(0, \frac{1}{I(\theta)})$.

If g is any function of θ s.t. $g'(\theta) \neq 0$.

$$\sqrt{n}(g(\hat{\theta}) - g(\theta)) \xrightarrow{d} N\left(0, \frac{[g'(\theta)]^2}{I(\theta)}\right) \quad \left(\text{By Delta theorem}\right).$$

when n is very large

the distribution of $\sqrt{n}(g(\hat{\theta}) - g(\theta))$ can be approximated by $N\left(0, \frac{[g'(\theta)]^2}{I(\theta)}\right)$

$$E\left[\sqrt{n}(g(\hat{\theta}) - g(\theta))\right] \approx 0 \Rightarrow E[g(\hat{\theta})] \approx g(\theta)$$

and $$Var\left(\sqrt{n}(g(\hat{\theta}) - g(\theta))\right) \approx \frac{[g'(\theta)]^2}{I(\theta)}$$

$$\Rightarrow Var(g(\hat{\theta})) \approx \frac{[g'(\theta)]^2}{n I(\theta)}$$

thus $g(\hat{\theta})$ is an asymptotically unbiased estimator of $g(\theta)$ whose variance asymptotically achieves the Cramer-Rao lower bound.

Asymptotic distribution of MLE is Normal when the regularity conditions hold.

- It is not always true.

<u>Example</u>: $X_1, \ldots, X_n \overset{iid}{\sim} U(0, \theta)$, $\theta > 0$.

$$f_\theta(x_1, \ldots, x_n) = \frac{1}{\theta^n} I(x_{(n)} < \theta)$$

If $\theta > X_{(n)}$ then $\frac{1}{\theta} < \frac{1}{X_{(n)}}$

Here the MLE is $X_{(n)}$.

Now, $\sqrt{n}(X_{(n)} - \theta) \not\xrightarrow{d} N(0, \frac{1}{I(\theta)})$

$n(X_{(n)} - \theta) \xrightarrow{d} Exp(\theta)$

<u>Bayes and Minimax estimator</u>

$X_1, \ldots, X_n \overset{iid}{\sim} f(x|\theta)$.

Assume a prior distribution on θ

Suppose the prior dist. is $\pi(\theta)$. There are objective and subjective choices of priors.

Posterior dist., denoted by $\pi(\theta|X_1, \ldots, X_n)$,

$$\pi(\theta|X_1, \ldots, X_n) = \frac{f(X_1, \ldots, X_n|\theta)\,\pi(\theta)}{\int f(X_1, \ldots, X_n|\theta)\,\pi(\theta)\,d\theta}$$

Example: $X_1, \dots, X_n \overset{iid}{\sim} Ber(p)$, $p \sim Beta(\alpha, \beta)$

$$\pi(p|x_1, \dots, x_n) \propto \left[\prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}\right] p^{\alpha-1}(1-p)^{\beta-1}$$

$$\propto p^{\sum_{i=1}^{n} x_i + \alpha - 1}(1-p)^{n - \sum_{i=1}^{n} x_i + \beta - 1}$$

Thus $p|x_1, \dots, x_n \sim Beta\left(\sum_{i=1}^{n} x_i + \alpha,\ n - \sum_{i=1}^{n} x_i + \beta\right)$

Example: $X_1, \dots, X_n \overset{iid}{\sim} Poi(\lambda)$, $\lambda \sim Gamma(\alpha, \beta)$

$$\pi(\lambda|x_1, \dots, x_n) \propto \left[\prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}\right] e^{-\lambda/\beta}\lambda^{\alpha-1}$$

$$\propto e^{-\lambda(n+\frac{1}{\beta})}\lambda^{\sum_{i=1}^{n} x_i + \alpha - 1}$$

$$\lambda|x_1, \dots, x_n \sim Gamma\left(\sum_{i=1}^{n} x_i + \alpha,\ \frac{1}{n+\frac{1}{\beta}}\right)$$

Definition (Conjugate family)

Let $\mathcal{F}$ denote the class of p.d.f.s or p.m.f.s $f(x|\theta)$. A class Π of prior distributions is a conjugate family for $\mathcal{F}$ if the posterior distribution is in the class Π for all $f \in \mathcal{F}$ and all priors in Π.

Now we will learn how to construct good estimators from Bayesian statistic. We will be using the "risk function" notion in Bayesian statistics.

Let $\delta(\underline{X})$ be an estimator of θ. The loss in estimating θ by $\delta(\underline{X})$ is represented by a function known as the loss function.

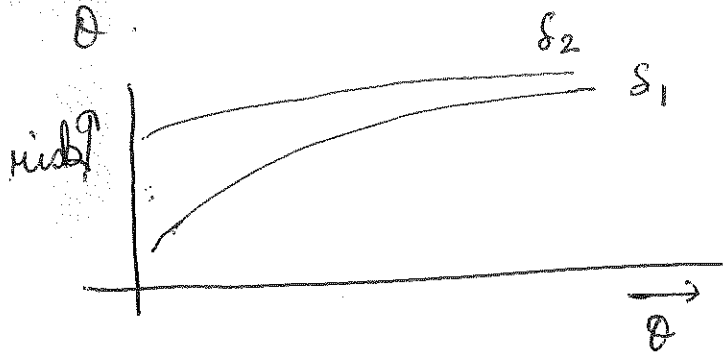
Mathematically we write it as $L(\theta, \delta)$

Define $R(\theta, \delta) = E_{\underline{X}|\theta}\left[L(\theta, \delta(\underline{X}))\right]$ is known as the risk function of δ.

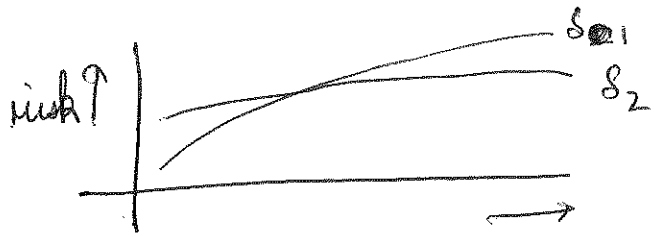
When $L(\theta, \delta(\underline{X})) = (\delta(\underline{X}) - \theta)^2$ then

$R(\theta, \delta(\underline{X})) = E_{\underline{X}|\theta}\left[(\delta(\underline{X}) - \theta)^2\right]$ is the MSE of $\delta(\underline{X})$.

Given two estimators δ_1, δ_2, we say δ_1 is better than δ_2 if $R(\theta, \delta_1) \leq R(\theta, \delta_2)$ for all θ and the inequality is strict for at least one θ.



δ_1 is better.



When two risks intersect each other, we do not know which estimator to choose. At that time, we have to propose a summary measure from the entire risk fn.

Different summary measures have been proposed. We will discuss the following two.

1. Average risk $= E_\theta E_{X|\theta}[L(\theta, \delta(\underline{X}))]$.

Minimizing this average risk will lead to the Bayes estimator.

2. Supremum risk $= \sup_\theta E_{\underline{X}|\theta}[L(\theta, \delta(\underline{X}))]$.

Minimizing the supremum risk will give rise to the Minimax estimator.

<u>Bayes estimator</u>

$$E_\theta E_{\underline{X}|\theta}[L(\theta, \delta(\underline{X}))] = E_{\underline{X},\theta}[L(\theta, \delta(\underline{X}))] = E_{\underline{X}} E_{\theta|\underline{X}}[L(\theta, \delta(\underline{X}))]$$

we have to minimize this Average risk w.r.t. $\delta(\underline{X})$.

we will find $\delta(\underline{X})$ that minimizes $E_{\theta|\underline{X}}[L(\theta, \delta(\underline{X}))]$ for all $\underline{X}$.

there are a few loss functions where it can be done easily.

<u>Squared error loss</u>

Let $L(\theta, \delta(\underline{X})) = (\delta(\underline{X}) - \theta)^2$

$$E_{\theta|\underline{X}}[(\delta(\underline{X}) - \theta)^2] = E_{\theta|\underline{X}}[(\delta(\underline{X}) - E[\theta|\underline{X}] + E[\theta|\underline{X}] - \theta)^2]$$

$$= E_{\theta|\underline{X}}[(\delta(\underline{X}) - E[\theta|\underline{X}])^2] + E_{\theta|\underline{X}}[(\theta - E[\theta|\underline{X}])^2]$$

$$E_{\theta|\underline{x}}\left[(\delta(\underline{x})-\theta)^2\right] \geq E_{\theta|\underline{x}}\left[(E[\theta|\underline{x}]-\theta)^2\right]$$

Thus $\delta(\underline{x}) = E[\theta|\underline{x}]$ is going to be the Bayes estimator.

$E_{\theta|\underline{x}}\left[L(\theta,\delta(\underline{x}))\right]$ you want to minimize w.r.t. $\delta(\underline{x})$

$$\frac{d}{d\delta} E_{\theta|\underline{x}}\left[L(\theta,\delta)\right] = 0$$

For $L(\theta,\delta) = (\theta-\delta)^2$

$$\frac{d}{d\delta} E_{\theta|\underline{x}}\left[(\theta-\delta)^2\right] = 0$$

$$\Rightarrow E_{\theta|\underline{x}}\left[2(\delta-\theta)\right] = 0$$

$$\Rightarrow E_{\theta|\underline{x}}[\delta] = E_{\theta|\underline{x}}[\theta]$$

$$\Rightarrow \delta(\underline{x}) = E_{\theta|\underline{x}}[\theta]$$

<u>Weighted squared error loss</u>

$$L(\theta,\delta(\underline{x})) = \omega(\theta)\left(\theta-\delta(\underline{x})\right)^2, \quad \omega(\theta) > 0.$$

$$\frac{d}{d\delta} E_{\theta|\underline{x}}\left[\omega(\theta)(\theta-\delta)^2\right] = 0$$

$$\Rightarrow E_{\theta|\underline{x}}\left[2\omega(\theta)(\delta-\theta)\right] = 0$$

$$\Rightarrow E_{\theta|\underline{x}}\left[\omega(\theta)\,\delta\right] = E_{\theta|\underline{x}}\left[\theta\,\omega(\theta)\right]$$

$$\Rightarrow \delta(\underline{x}) = \frac{E_{\theta|\underline{x}}\left[\theta\,\omega(\theta)\right]}{E_{\theta|\underline{x}}\left[\omega(\theta)\right]}$$

Example: $X_1, \dots, X_n \overset{iid}{\sim} Ber(p)$. $p \sim Beta(\alpha, \beta)$

$p | X_1, \dots, X_n \sim Beta\left(\alpha + \sum_{i=1}^n x_i, \; n - \sum_{i=1}^n x_i + \beta\right)$

under squared error loss $L(\theta, \delta(\underline{x})) = (\theta - \delta(\underline{x}))^2$

What is the Bayes estimator?

Bayes estimator $E[p | x_1, \dots, x_n] = \dfrac{\alpha + \sum_{i=1}^n x_i}{\alpha + \sum_{i=1}^n x_i + n - \sum_{i=1}^n x_i + \beta}$

$$= \frac{\alpha + \sum_{i=1}^n x_i}{\alpha + \beta + n}.$$

$$E\left[E[p | x_1, \dots, x_n]\right] = \frac{\alpha + np}{\alpha + \beta + n} \neq p$$

this is a biased estimator of p.

Theorem: No unbiased estimator $\delta(\underline{x})$ of θ can be a Bayes estimator under squared error loss unless $E_\theta E_{\underline{x}|\theta}\left[(\theta - \delta(\underline{x}))^2\right] = 0$

Pf:- Let $\delta(\underline{x})$ be an unbiased estimator which is also a Bayes estimator.

$E[\delta(\underline{x}) | \theta] = \theta$ for all θ and $E[\theta | \underline{x}] = \delta(\underline{x})$

Now,

$$E_\theta E_{\underline{x}|\theta}\left[\delta(\underline{x})\theta\right] = E_\theta\left[\theta E_{\underline{x}|\theta}[\delta(\underline{x})]\right] = E_\theta[\theta^2]$$

$$E_{\underline{x}} E_{\theta|\underline{x}}\left[\delta(\underline{x})\theta\right] = E_{\underline{x}}\left[\delta(\underline{x}) E_{\theta|\underline{x}}[\theta]\right] = E_{\underline{x}}\left[\delta(\underline{x})^2\right]$$

$$E_\theta E_{\underline{X}|\theta}[\delta(\underline{X})^2] = E_\theta E_{\underline{X}|\theta}[\delta(\underline{X})\theta] = E_\theta E_{\underline{X}|\theta}[\theta^2] \quad \cdots (*)$$

Hence,

$$E_{\theta,\underline{X}}[(\theta-\delta(\underline{X}))^2] = E_{\theta,\underline{X}}[\theta^2 - 2\delta(\underline{X})\theta + \delta(\underline{X})^2]$$

By $(*)$ $\qquad = 0$

Sometimes you can show that the Bayes estimator is a convex combination of the prior mean and data mean.

<u>Example</u>: $X_1, \ldots, X_n \overset{iid}{\sim} Ber(p)$, $\quad p \sim Beta(\alpha, \beta)$

$$\Rightarrow \quad E[p \mid X_1, \ldots, X_n] = \frac{\sum_{i=1}^n X_i + \alpha}{\alpha+\beta+n}$$

$$= \underbrace{\frac{\sum_{i=1}^n X_i}{n}}_{\text{data mean}} \cdot \underbrace{\frac{n}{(\alpha+\beta+n)}}_{\omega_1} + \underbrace{\left(\frac{\alpha}{\alpha+\beta}\right)}_{\text{prior mean}} \underbrace{\frac{(\alpha+\beta)}{(\alpha+\beta+n)}}_{\omega_2}$$

$\omega_1 + \omega_2 = 1$

<u>Midterm</u>

1. $X_1, \ldots, X_{2n} \overset{iid}{\sim} Ber(p)$, $\qquad S_n = \sum_{i=1}^n X_{2i} X_{2i-1}$

$\sqrt{n}\left(\frac{S_n}{n} - p^2\right) \qquad E\left[\frac{S_n}{n}\right] = p^2$

$$Var(X_{2i} X_{2i-1}) = E[X_{2i}^2 X_{2i-1}^2] - \{E[X_{2i}]E[X_{2i-1}]\}^2$$

$$= E[X_{2i}^2]\, E[X_{2i-1}^2] - p^4$$

$$= \{p(1-p) + p^2\}\{p(1-p) + p^2\} - p^4$$

$\text{var}(X_{2i} X_{2i-1}) = p^2 - p^4 = p^2(1-p^2)$

CLT tells us $\sqrt{n}\left(\frac{S_n}{n} - p^2\right) \xrightarrow{d} N(0, p^2(1-p^2))$

2. $X = \begin{cases} -1 & \text{w.p. } \theta/2 \\ 0 & \text{w.p. } 1-\theta \\ 1 & \text{w.p. } \theta/2 \end{cases}$

$|X_i| = 1$ w.p. θ

$= 0$ w.p. $1-\theta$

$\Rightarrow |X_1|, \ldots, |X_n| \overset{iid}{\sim} \text{Ber}(\theta)$.

Now the rest follows from classes.

θ^2 that was $\dfrac{\sum_{i=1}^{n} |X_i| \left(\sum_{i=1}^{n} |X_i| - 1\right)}{n(n-1)}$

3. $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$,

$(\bar{X}, S^2)$ is a complete sufficient stat.

$E[\bar{X}] = \mu$ thus By Lehman-Scheffe theorem

$\bar{X}$ is the UMVUE for μ.

$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ $= \text{Gamma}\left(\frac{n-1}{2}, 2\right)$

$\Rightarrow \frac{\sigma^2}{(n-1)S^2} \sim \text{Inverse-Gamma}\left(\frac{n-1}{2}, 2\right)$

$\Rightarrow E\left[\frac{\sigma^2}{(n-1)S^2}\right] = \frac{1}{2\left(\frac{n-1}{2} - 1\right)} = \frac{1}{n-3}$

$\Rightarrow E\left[\frac{n-3}{(n-1)S^2}\right] = \frac{1}{\sigma^2}$ this is a function of the complete sufficient stat.

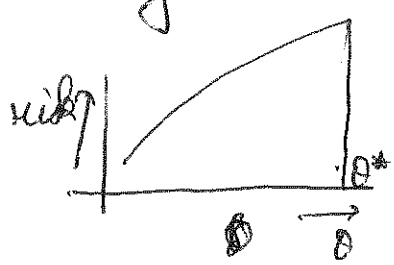
Recap:

① Risk function $E_{\underline{X}|\theta}\left[L(\theta, \delta(\underline{X}))\right]$ of an estimator $\delta(\underline{X})$.

② $E_\theta E_{\underline{X}|\theta}\left[L(\theta, \delta(\underline{X}))\right] \longrightarrow$ an estimator $\delta(\underline{X})$ that minimizes this average risk fn. is called the Bayes estimator

$\sup_\theta E_{\underline{X}|\theta}\left[L(\theta, \delta(\underline{X}))\right]$. The Minimax estimator minimizes $\sup_\theta E_{\underline{X}|\theta}\left[L(\theta, \delta(\underline{X}))\right]$. Thus minimax estimator is going to protect us from the worst case scenario.

Computing/finding minimax estimator is difficult. However, we can use other tricks to compute minimax estimators. If the prior distribution of θ gives high probability to those values, which will increase the risk, then intuitively a minimax estimator is a Bayes estimator.



If the prior only has mass at θ^* then the

$$E_\theta E_{\underline{X}|\theta}\left[L(\theta, \delta(\underline{X}))\right] = \sup_\theta E_{\underline{X}|\theta}\left[L(\theta, \delta(\underline{X}))\right]$$

Definition (Least favorable distribution)

A prior distribution $\pi(\theta)$ is known to be a least favorable prior if $E_\theta E_{\underline{X}|\theta}[L(\theta, \delta_\pi(\underline{X}))]$ $\geq E_\theta E_{\underline{X}|\theta}[L(\theta, \delta_{\pi'}(\underline{X}))]$ for all ~~priors~~ prior distributions $\pi'(\theta)$ on θ. Here δ_π and $\delta_{\pi'}$ are Bayes estimators w.r.t. priors π and π' respectively.

So, $\pi(\theta)$ is such that ~~the~~ it increases the Bayes risk for the Bayes estimator w.r.t. any other prior.

<u>Result</u>: If $\pi(\theta)$ be a prior distribution for which $\int E_{\underline{X}|\theta}[L(\theta, \delta_\pi)]\, d\pi(\theta) = \sup_\theta E_{\underline{X}|\theta}[L(\theta, \delta_\pi)]$ ~~then~~ where δ_π is the Bayes estimator. Then

(a) δ_π is minimax estimator

(b) π is least favorable distribution

<u>proof</u>:- a) For any $\delta(\underline{X})$,

$$\sup_\theta E_{\underline{X}|\theta}[L(\theta, \delta(\underline{X}))] \geq E_\theta E_{\underline{X}|\theta}[L(\theta, \delta(\underline{X}))]$$

$$\geq E_\theta E_{\underline{X}|\theta}[L(\theta, \delta_\pi(\underline{X}))] \underset{\text{by assumption}}{=} \sup_\theta E_{\underline{X}|\theta}[L(\theta, \delta_\pi(\underline{X}))]$$

Thus $\delta_\pi(\underline{X})$ is a minimax estimator

(b) Note that,

$$E_\theta E_{\underline{x}|\theta}\left[L(\theta, \delta_\pi)\right] \overset{\text{by assumption}}{=} \sup_\theta E_{\underline{x}|\theta}\left[L(\theta, \delta_\pi)\right]$$

$$\geq \int E_{\underline{x}|\theta}\left[L(\theta, \delta_\pi)\right] \pi'(\theta)\, d\theta$$, for any other prior dist. $\pi'(\theta)$ on θ

$$E_{\underline{x}|\theta}\left[L(\theta, \delta_\pi)\right] \leq \sup_\theta E_{\underline{x}|\theta}\left[L(\theta, \delta_\pi)\right]$$

$$\int E_{\underline{x}|\theta}\left[L(\theta, \delta_\pi)\right] \pi'(\theta)\, d\theta \leq \sup_\theta E_{\underline{x}|\theta}\left[L(\theta, \delta_\pi)\right] \underbrace{\int \pi'(\theta)\, d\theta}_{=1}$$

$$\geq \int E_{\underline{x}|\theta}\left[L(\theta, \delta_{\pi'})\right] \pi'(\theta)\, d\theta$$

$$= E_\theta E_{\underline{x}|\theta}\left[L(\theta, \delta_{\pi'})\right]$$

Thus, $E_\theta E_{\underline{x}|\theta}\left[L(\theta, \delta_\pi)\right] \geq E_\theta E_{\underline{x}|\theta}\left[L(\theta, \delta_{\pi'})\right]$ which implies that $\pi^*(\theta)$ is the least favorable distribution.

To use this result, we need to have a prior dist. for which

$$\int E_{\underline{x}|\theta}\left[L(\theta, \delta_\pi)\right] \pi(\theta)\, d\theta = \sup_\theta E_{\underline{x}|\theta}\left[L(\theta, \delta_\pi)\right]$$

then we can find the Bayes estimator w.r.t. that prior. Our result says that this Bayes estimator will be the minimax estimator

If the risk fn. $E_{\underline{x}|\theta}[L(\theta, \delta_\pi)]$ is const. then average is going to be the supremum.

One algorithm.

Step 1: Take a prior dist of your choice.

Step 2: ~~Compute the risk fn. w.r.t.~~ Compute the risk fn. of the Bayes estimator w.r.t. that prior.

Step 3: Try to find prior hyperparameters in such a way that the risk fn. of the Bayes estimator becomes const.

Example: $x_1, \ldots, x_n \overset{iid}{\sim} Ber(p)$, $p \sim Beta(\alpha, \beta)$.

$$\delta_\pi(\underline{x}) = \frac{\sum_{i=1}^n x_i + \alpha}{n+\alpha+\beta}.$$

$$E_{\underline{x}|p}\left[(\delta_\pi(\underline{x}) - p)^2\right] = E_{\underline{x}|p}\left[\left(\frac{\sum_{i=1}^n x_i + \alpha}{n+\alpha+\beta} - p\right)^2\right]$$

$$= Var_{\underline{x}|p}\left(\frac{\sum_{i=1}^n x_i + \alpha}{n+\alpha+\beta}\right) + \left[E_{\underline{x}|p}\left(\frac{\sum_{i=1}^n x_i + \alpha}{n+\alpha+\beta}\right) - p\right]^2$$

$$= \frac{np(1-p)}{(n+\alpha+\beta)^2} + \left[\frac{np+\alpha}{n+\alpha+\beta} - p\right]^2$$

$$= \frac{np(1-p)}{(n+\alpha+\beta)^2} + \frac{[\alpha - p\alpha - p\beta]^2}{(n+\alpha+\beta)^2}$$

$$= \frac{1}{(n+\alpha+\beta)^2}\left[\alpha^2 + \{n - 2\alpha(\alpha+\beta)\}p + \{(\alpha+\beta)^2 - n\}p^2\right]$$

This to be a const. (not a function of p) we must have

$$n = 2\alpha(\alpha+\beta) \quad \text{--- (1)}$$

$$\sqrt{n} = (\alpha+\beta) \quad \text{--- (2)}$$

solving them we obtain $\alpha = \frac{\sqrt{n}}{2}$, $\beta = \frac{\sqrt{n}}{2}$.

Thus by our earlier thm., $\text{Beta}\left(\frac{\sqrt{n}}{2}, \frac{\sqrt{n}}{2}\right)$ is a least favorable dist. and

$$\frac{\sum_{i=1}^{n} x_i + \frac{\sqrt{n}}{2}}{n+\sqrt{n}}$$ is a minimax estimator.

Testing of Hypothesis

Statistical hypothesis testing is all about

1. Beginning with a tentative idea about the unknown parameter.
2. Want to test the validity of this tentative idea based on sample. Existing tentative idea, status quo: H_0 (null hypothesis), new idea: H_1 (alternative hypothesis).
3. We begin by assuming that the null hypothesis is true. Only when there is an overwhelming evidence contradicting the null do we reject it in favor of the alternative.

	H_0 is true	H_0 is false
Do not reject H_0	Correct	Incorrect
Reject H_0	Incorrect	Correct

Type 1 error $= P(\text{reject } H_0 \mid H_0 \text{ is true})$

Type 2 error $= P(\text{do not reject } H_0 \mid H_0 \text{ is false})$

Type 1 error is also known as the level (size) of the test.

On the other hand, 1 − Type 2 error

$= P(\text{Rejecting } H_0 \mid H_0 \text{ is false}) =$ power of the test.

$0 \le \text{level} \le 1, \quad 0 \le \text{power} \le 1$

Ideally, we would like to minimize both type 1 and type 2 error. But it turns out that it is not possible to simultaneously minimize them. So, we fix level at a pre-specified value and find a test that maximizes power.

<u>Parametric tests</u>: $X_1, \ldots, X_n \overset{iid}{\sim} f(x \mid \theta)$. We test $H_0: \theta \in \Theta_0$ vs. $H_1: \theta \in \Theta_1$. If Θ_0 is singleton we will call it a simple null hypothesis. Otherwise we will call it a composite null hypothesis

Let $\mathcal{R} = \left\{ \underline{x} \in \mathcal{X} \;\middle|\; \begin{array}{l}\text{the null hypothesis is} \\ \text{rejected for } \underline{x}\end{array} \right\}$.

$\mathcal{R}$ is called the rejection region or critical region of the test.

$\phi(\underline{x})$ = prob. of rejecting H_0 when $\underline{x}$ is observed.

Power function of a test is given by

$$\beta(\theta) = \int \phi(\underline{x}) f(\underline{x}|\theta)\, d\underline{x}$$

$$= \int P(\text{rejecting } H_0 \mid \underline{x} \text{ is observed})\, P(\underline{x} \text{ is observed} \mid \theta \text{ is the truth})\, d\underline{x}$$

intuitively

$$= P(\text{reject } H_0 \mid \theta \text{ is the truth}).$$

If level of a test is α and the power fn. of that test is given by $\beta(\theta)$ and ~~the~~ hypothesis is $H_0: \theta \in \Theta_0$ vs. $H_1: \theta \in \Theta_1$, then $\beta(\theta) \le \alpha$, $\theta \in \Theta_0$