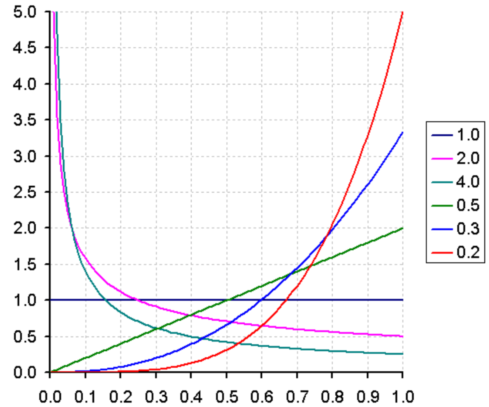
STAT 410 **Examples for 10/16/2020 (1)** Fall 2020
A. Stepanov

p.m.f. or p.d.f. $f(x;\theta)$, $\theta \in \Omega$. Ω – parameter space.

1. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the distribution with probability density function

$$f(x;\theta) = \begin{cases} \dfrac{1}{\theta} \cdot x^{\frac{1}{\theta}-1} & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$0 < \theta < \infty$.



a) Obtain the maximum likelihood estimator of θ, $\hat{\theta}$.

Idea: Out of all possible values of the parameter θ, choose the one that gives you the best chance, the maximum likelihood to obtain a data set just like the one we have.

Likelihood function:

$$L(\theta) = L(\theta; x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i;\theta) = f(x_1;\theta) \cdot \ldots \cdot f(x_n;\theta)$$

It is often easier to consider $\ln L(\theta) = \sum_{i=1}^{n} \ln f(x_i;\theta)$.

Maximum Likelihood Estimator: $\hat{\theta} = \arg\max L(\theta) = \arg\max \ln L(\theta)$.

Likelihood function:

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta) = \frac{1}{\theta^n} \cdot \left(\prod_{i=1}^{n} x_i \right)^{\frac{1}{\theta}-1}.$$

$$\ln L(\theta) = -n \cdot \ln\theta + \left(\frac{1}{\theta} - 1\right) \cdot \sum_{i=1}^{n} \ln x_i .$$

$$\frac{d}{d\theta}\left(\ln L\left(\hat{\theta}\right)\right) = -\frac{n}{\hat{\theta}} - \frac{1}{\hat{\theta}^2} \cdot \sum_{i=1}^{n} \ln x_i = 0. \qquad \Rightarrow \qquad \hat{\theta} = -\frac{1}{n} \cdot \sum_{i=1}^{n} \ln X_i .$$

b) Obtain the method of moments estimator of θ, $\tilde{\theta}$.

Idea: Out of all possible values of the parameter θ, choose the one that makes the sample mean equal to the population mean.

Method of Moments:

$E(X) = h(\theta)$. Set $\overline{X} = h(\tilde{\theta})$. Solve for $\tilde{\theta}$.

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x;\theta)\, dx = \int_{0}^{1} x \cdot \frac{1}{\theta} \cdot x^{\frac{1}{\theta}-1}\, dx = \frac{1}{\theta+1}.$$

$$\overline{X} = \frac{1}{1+\tilde{\theta}}. \qquad \Rightarrow \qquad \tilde{\theta} = \frac{1-\overline{X}}{\overline{X}} = \frac{1}{\overline{X}} - 1.$$

c) Suppose $n = 3$, and $x_1 = 0.2$, $x_2 = 0.3$, $x_3 = 0.5$. Compute the values of the method of moments estimate and the maximum likelihood estimate for θ.

$$\overline{X} = \frac{0.2+0.3+0.5}{3} = \frac{1}{3}. \qquad \tilde{\theta} = \frac{1-\overline{X}}{\overline{X}} = \frac{1-1/3}{1/3} = \mathbf{2}.$$

$$\hat{\theta} = -\frac{1}{n} \cdot \sum_{i=1}^{n} \ln X_i = -\frac{1}{3} \cdot (\ln 0.2 + \ln 0.3 + \ln 0.5) = -\frac{1}{3} \cdot \ln 0.03 \approx \mathbf{1.16885}.$$

d) What is the probability distribution of $W = -\ln X$?

Why $W = -\ln X$? Because the maximum likelihood estimator $\hat{\theta}$ is made out of them. If we want to know more about $\hat{\theta}$, we may want to know more about the distribution of $W = -\ln X$.

$$F_X(x) = x^{1/\theta}, \qquad 0 < x < 1.$$

Then $F_W(w) = P(W \le w) = P(X \ge e^{-w})$

$$= 1 - F_X(e^{-w}) = 1 - e^{-w/\theta}, \qquad w > 0.$$

$\Rightarrow$ W has Exponential(θ) = Gamma$(\alpha = 1, \theta)$ distribution.

Recall: $X \sim$ Gamma(α_1, θ), $Y \sim$ Gamma(α_2, θ), X and Y are independent.

$\Rightarrow$ $X + Y \sim$ Gamma$(\alpha_1 + \alpha_2, \theta)$.

e) Suppose $n = 3$ and $\theta = 1.25$. Find $P(-\sum_{i=1}^{3} \ln X_i > 3.5)$.

$-\sum_{i=1}^{3} \ln X_i = \sum_{i=1}^{3} W_i$ has a Gamma distribution with $\alpha = 3$ and $\theta = 1.25$.

Recall: If T has a Gamma$(\alpha, \theta = {}^1\!/\!_\lambda)$ distribution, where α is an integer,

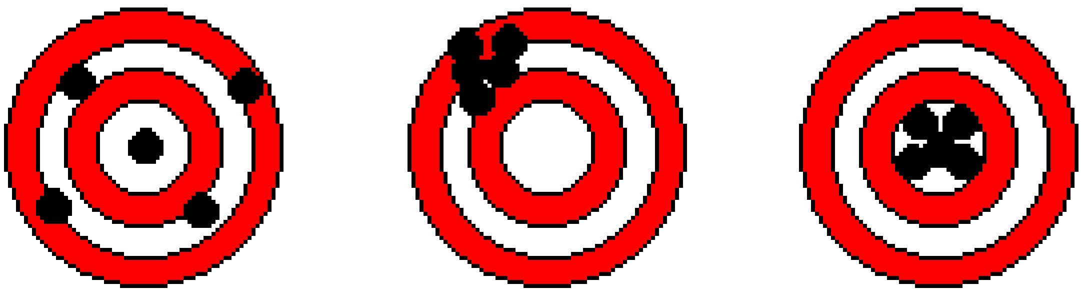
then $F_T(t) = P(T \le t) = P(X_t \ge \alpha)$ and $P(T > t) = P(X_t \le \alpha - 1)$,

where X_t has a Poisson$(\lambda t = {}^t\!/\!_\theta)$ distribution.

$$P(-\sum_{i=1}^{3} \ln X_i > 3.5) = P(T_3 > 3.5) = P(X_{3.5} \le 3 - 1)$$

$$= P(\text{Poisson}(\frac{3.5}{1.25}) \le 2) = P(\text{Poisson}(2.8) \le 2) = \mathbf{0.469}.$$

Def An estimator $\hat{\theta}$ is said to be **unbiased for θ** if $E(\hat{\theta}) = \theta$ for all θ.



accurate but imprecise	inaccurate but precise	accurate and precise
unbiased, large variance	biased, small variance	unbiased, small variance

Obviously, ③ is better than ① or ②. How do we compare ① and ② ?

Def For an estimator $\hat{\theta}$ of θ, define the **Mean Squared Error** of $\hat{\theta}$ by

$$\text{MSE}(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right].$$

$$\begin{aligned}
E\left[(\hat{\theta} - \theta)^2\right] &= E\left[\left(\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta\right)^2\right] \\
&= E\left[\left(\hat{\theta} - E(\hat{\theta})\right)^2\right] + E\left[\left(E(\hat{\theta}) - \theta\right)^2\right] \\
&\qquad + 2E\left[\left(\hat{\theta} - E(\hat{\theta})\right)\left(E(\hat{\theta}) - \theta\right)\right] \\
&= \text{Var}(\hat{\theta}) + \left(E(\hat{\theta}) - \theta\right)^2 + 2E\left[\left(\hat{\theta} - E(\hat{\theta})\right)\right]\left(E(\hat{\theta}) - \theta\right) \\
&= \text{Var}(\hat{\theta}) + \left(E(\hat{\theta}) - \theta\right)^2 = \text{Var}(\hat{\theta}) + \left(\text{bias}(\hat{\theta})\right)^2.
\end{aligned}$$

If $\hat{\theta}$ is an unbiased estimator for θ, $\quad \text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta})$.

f) Is $\hat{\theta}$ unbiased for θ? That is, does $\text{E}(\hat{\theta})$ equal θ?

$\hat{\theta} = \overline{\text{W}}$. $\quad \text{E}(\hat{\theta}) = \text{E}(\overline{\text{W}}) = \text{E}(\text{W}) = \theta$.

That is, $\hat{\theta}$ is an unbiased estimator for θ.

OR

$$\text{E}(\ln \text{X}) = \int_{-\infty}^{\infty} \ln x \cdot f(x;\theta)\, dx = \int_0^1 \ln x \cdot \frac{1}{\theta} \cdot x^{\frac{1}{\theta}-1}\, dx.$$

By parts: $\quad u = \ln x, \quad du = \frac{1}{x}\, dx,$

$$dv = \frac{1}{\theta} \cdot x^{\frac{1}{\theta}-1}\, dx, \qquad v = x^{1/\theta}.$$

$$\text{E}(\ln \text{X}) = \int_0^1 \ln x \cdot \frac{1}{\theta} \cdot x^{\frac{1}{\theta}-1}\, dx = \left(\ln x \cdot x^{1/\theta} \right) \Big|_0^1 - \int_0^1 \frac{1}{x} \cdot x^{1/\theta}\, dx$$

$$= -\int_0^1 x^{\frac{1}{\theta}-1}\, dx = -\left(\frac{1}{1/\theta} \cdot x^{1/\theta} \right) \Big|_0^1 = -\theta.$$

Therefore, $\quad \text{E}\left(\hat{\theta}\right) = -\frac{1}{n} \cdot \sum_{i=1}^{n} \text{E}(\ln \text{X}_i) = -\frac{1}{n} \cdot \sum_{i=1}^{n} (-\theta) = \theta,$

That is, $\hat{\theta}$ is an unbiased estimator for θ.

g) Find $\mathrm{MSE}(\hat{\theta}) = \mathrm{E}\left[(\hat{\theta}-\theta)^2\right] = \left(\mathrm{bias}(\hat{\theta})\right)^2 + \mathrm{Var}(\hat{\theta})$.

$$\mathrm{Var}(\hat{\theta}) = \mathrm{Var}(\overline{\mathrm{W}}) = \frac{\mathrm{Var}(\mathrm{W})}{n} = \frac{\theta^2}{n}.$$

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{Var}(\hat{\theta}) + \left(\mathrm{bias}(\hat{\theta})\right)^2 = \frac{\theta^2}{n} + 0 = \frac{\theta^2}{n} \to 0 \quad \text{as } n \to \infty.$$

$\hat{\theta}$ is a random variable, different samples may (and likely will) give us different values of $\hat{\theta}$.

Part (g) suggests that if n is large, then $\hat{\theta}$ will be close to θ with high probability.

This is the most we can possibly hope for. It would be unreasonable to hope for $\hat{\theta}$ to be equal to θ. All we can hope for is that $\hat{\theta}$ will be close to θ with high probability.

$\hat{\theta}$ is an unbiased estimator for θ. For large n, $\mathrm{Var}(\hat{\theta})$ is small (just like ③).

$\hat{\theta}$ is a wonderful estimator for θ!

In general, MLE estimators tend to be "better" than MoM estimators
(we will find out why later).

h) Is $\tilde{\theta}$ unbiased for θ? That is, does $\mathrm{E}(\tilde{\theta})$ equal θ?

$$\tilde{\theta} = g(\overline{\mathrm{X}}), \quad \text{where } g(x) = \frac{1-x}{x} = \frac{1}{x} - 1, \quad 0 < x < 1.$$

$$\mathrm{E}(\overline{\mathrm{X}}) = \mu = \frac{1}{1+\theta}. \qquad g(\mu) = \theta. \qquad \mathrm{E}(g(\overline{\mathrm{X}})) = ???$$

IF $g(x) = a\,x + b$ is a linear function, then

$$\mathrm{E}(g(\mathrm{X})) = g(\mathrm{E}(\mathrm{X})). \qquad \mathrm{E}(g(a\,\mathrm{X} + b)) = a\,\mu_{\mathrm{X}} + b.$$

In general, $\mathrm{E}(g(\mathrm{X})) \neq g(\mathrm{E}(\mathrm{X}))$.

= =

Jensen's Inequality: (Theorem 1.10.5)

If g is convex on an open interval I and X is a random variable whose support is contained in I and has finite expectation, then

$$\mathrm{E}[g(\mathrm{X})] \ge g[\mathrm{E}(\mathrm{X})].$$

If g is strictly convex then the inequality is strict, unless X is a constant random variable.

$\Rightarrow \quad \mathrm{E}(\mathrm{X}^2) \ge [\mathrm{E}(\mathrm{X})]^2 \quad \Leftrightarrow \quad \mathrm{Var}(\mathrm{X}) \ge 0$

$\Rightarrow \quad \mathrm{E}(e^{t\mathrm{X}}) \ge e^{t\mathrm{E}(\mathrm{X})} \quad \Rightarrow \quad \mathrm{M}_{\mathrm{X}}(t) \ge e^{t\mu}$

$\Rightarrow \quad \mathrm{E}\left(\dfrac{1}{\mathrm{X}}\right) \ge \dfrac{1}{\mathrm{E}(\mathrm{X})}$ for a positive random variable X

$\Rightarrow \quad \mathrm{E}[\mathrm{X}^3] \ge [\mathrm{E}(\mathrm{X})]^3$ for a non-negative random variable X

$\Rightarrow \quad \mathrm{E}[\ln \mathrm{X}] \le \ln \mathrm{E}(\mathrm{X})$ for a positive random variable X

$\Rightarrow \quad \mathrm{E}\left(\sqrt{\mathrm{X}}\right) \le \sqrt{\mathrm{E}(\mathrm{X})}$ for a non-negative random variable X

= =

$g(x) = \dfrac{1-x}{x} = \dfrac{1}{x} - 1, \quad 0 < x < 1. \qquad g''(x) = \dfrac{2}{x^3} > 0, \quad 0 < x < 1.$

Since $g(x) = \dfrac{1-x}{x} = \dfrac{1}{x} - 1$, $0 < x < 1$, is strictly convex, and $\overline{\mathrm{X}}$ is not a constant random variable, by Jensen's Inequality (Theorem 1.10.5),

$$\mathrm{E}(\tilde{\theta}) = \mathrm{E}(g(\overline{\mathrm{X}})) > g(\mathrm{E}(\overline{\mathrm{X}})) = g(\mu) = \theta.$$

$\tilde{\theta}$ is NOT an unbiased estimator for θ.

2. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a Geometric (p) distribution (the number of independent trials until the first "success"). That is,

$$P(X_1 = k) = (1-p)^{k-1}\, p, \quad k = 1, 2, 3, \ldots .$$

a) Obtain the method of moments estimator of p, $\tilde{p}$.

$E(X) = {}^1\!/_p$. $\quad \overline{X} = {}^1\!/_{\tilde{p}}$ so $\tilde{p} = {}^1\!/_{\overline{X}} = {}^n\!/_{\sum_{i=1}^n X_i}$.

b) Obtain the maximum likelihood estimator of p, $\hat{p}$.

$$L(p) = (1-p)^{\sum_{i=1}^n X_i - n}\, p^n$$

$$\ln L(p) = \left(\sum_{i=1}^n X_i - n\right) \ln(1-p) + n \ln p$$

$$\frac{d}{dp} \ln L(p) = \frac{n}{p} - \frac{\sum_{i=1}^n X_i - n}{1-p} = \frac{n - p\sum_{i=1}^n X_i}{p(1-p)}$$

$$\frac{d}{dp} \ln L(\hat{p}) = 0 \qquad \Rightarrow \qquad \hat{p} = {}^n\!/_{\sum_{i=1}^n X_i} = {}^1\!/_{\overline{X}}.$$

$\hat{p} = \tilde{p}$ equals the number of successes, n, divided by the number of Bernoulli trials, $\sum_{i=1}^n X_i$;

c) Is $\hat{p}$ an unbiased estimator for p?

Since $g(x) = {}^1\!/_x$, $x \ge 1$, is strictly convex, and $\overline{X}$ is not a constant random variable, by Jensen's Inequality (Theorem 1.10.5),

$$E(\hat{p}) = E(g(\overline{X})) > g(E(\overline{X})) = g({}^1\!/_p) = p.$$

$\hat{p}$ is NOT an unbiased estimator for p.

3. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a population with mean μ and variance σ^2. Show that the sample mean $\overline{X}$ and the sample variance S^2 are unbiased for μ and σ^2, respectively.

Random Sample: $X_1, X_2, \ldots, X_n$ are i.i.d. (independent, identically distributed).

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}, \qquad S^2 = \frac{1}{n-1}\sum\left(X_i - \overline{X}\right)^2.$$

$$E(X_1 + X_2 + \ldots + X_n) = n \cdot \mu. \qquad \Rightarrow \qquad E(\overline{X}) = \mu. \qquad \checkmark$$

$$E(X^2) = \mathrm{Var}(X) + [E(X)]^2 = \mu^2 + \sigma^2.$$

$$\mathrm{Var}(X_1 + X_2 + \ldots + X_n) = n \cdot \sigma^2. \qquad \Rightarrow \qquad \mathrm{Var}(\overline{X}) = \sigma^2/n.$$

$$E\left(\left(\overline{X}\right)^2\right) = \mathrm{Var}(\overline{X}) + [E(\overline{X})]^2 = \mu^2 + \frac{\sigma^2}{n}.$$

$$S^2 = \frac{1}{n-1}\sum\left(X_i - \overline{X}\right)^2 = \frac{1}{n-1}\left[\sum X_i^2 - n \cdot \left(\overline{X}\right)^2\right].$$

$$E(S^2) = \frac{1}{n-1}\left[\sum E\left(X_i^2\right) - n \cdot E\left(\left(\overline{X}\right)^2\right)\right]$$

$$= \frac{1}{n-1}\left[n \cdot \left(\mu^2 + \sigma^2\right) - n \cdot \left(\mu^2 + \frac{\sigma^2}{n}\right)\right] = \sigma^2. \qquad \checkmark$$

4. a) Let S^2 be the sample variance of a random sample from a distribution with variance $\sigma^2 > 0$. Since $E(S^2) = \sigma^2$, why isn't $E(S) = \sigma$?

Hint: Use Jensen's inequality to show that $E(S) < \sigma$.

$g(x) = x^2$ is strictly convex. By Jensen's Inequality,

$\sigma^2 = E(S^2) = E[g(S)] > g[E(S)] = [E(S)]^2.$

Therefore, $\qquad \sigma > E(S).$

OR

$E(S^2) = \sigma^2.$

$g(x) = -\sqrt{x}$, $x > 0$, is strictly convex. By Jensen's Inequality,

$-E(S) = E[g(S^2)] > g[E(S^2)] = -\sigma.$

Therefore, $\qquad E(S) < \sigma.$

b) Suppose that the sample is drawn from a $N(\mu, \sigma^2)$ distribution. Recall that $(n-1)S^2/\sigma^2$ has a $\chi^2(n-1)$ distribution. Use Theorem 3.3.1 to determine an unbiased estimator of σ.

Hint: That is, find b such that $E(bS) = \sigma$. $\qquad$ (b would depend on n)

Theorem 3.3.1. Let X have a $\chi^2(r)$ distribution. If $k > -r/2$, then $E(X^k)$ exists and it is given by

$$E(X^k) = \frac{2^k \Gamma\left(\frac{r}{2} + k\right)}{\Gamma\left(\frac{r}{2}\right)}.$$

By Theorem 3.3.1, if $r = n - 1$ and $k = {}^{1}\!/_{2}$, then

$$\mathrm{E}\left(\frac{\sqrt{n-1}\,\mathrm{S}}{\sigma} \right) = \frac{\sqrt{2}\,\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}.$$

Therefore,

$$\mathrm{E}\left(\frac{\sqrt{n-1}\,\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{2}\,\Gamma\left(\frac{n}{2}\right)} \cdot \mathrm{S} \right) = \sigma,$$

and $\dfrac{\sqrt{n-1}\,\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{2}\,\Gamma\left(\frac{n}{2}\right)} \cdot \mathrm{S}$ is unbiased for σ.

$n = 5$ $\qquad b = \dfrac{\sqrt{4}\,\Gamma\left(\frac{4}{2}\right)}{\sqrt{2}\,\Gamma\left(\frac{5}{2}\right)} = \dfrac{2 \cdot 1}{\sqrt{2} \cdot \left(\frac{3}{2}\right)\left(\frac{1}{2}\right)\sqrt{\pi}} = \dfrac{8}{3\sqrt{2\pi}} \approx 1.063846.$

$n = 6$ $\qquad b = \dfrac{\sqrt{5}\,\Gamma\left(\frac{5}{2}\right)}{\sqrt{2}\,\Gamma\left(\frac{6}{2}\right)} = \dfrac{\sqrt{5} \cdot \left(\frac{3}{2}\right)\left(\frac{1}{2}\right)\sqrt{\pi}}{\sqrt{2} \cdot 2} = \dfrac{3\sqrt{5\pi}}{8\sqrt{2}} \approx 1.050936.$

c) Suppose that the sample is drawn from a $\mathrm{N}(\mu, \sigma^2)$ distribution. Which value of c minimizes $\mathrm{MSE}(c\,\mathrm{S}^2) = \mathrm{E}[(c\,\mathrm{S}^2 - \sigma^2)^2]$? ($c$ would depend on n)

Hint: Recall that $\mathrm{E}(\chi^2(r)) = r$, $\mathrm{Var}(\chi^2(r)) = 2r$.

$$\mathrm{E}(\mathrm{S}^2) = \frac{\sigma^2}{n-1} \cdot \mathrm{E}\left(\frac{(n-1) \cdot \mathrm{S}^2}{\sigma^2} \right) = \frac{\sigma^2}{n-1} \cdot (n-1) = \sigma^2.$$

$$\mathrm{Var}(\mathrm{S}^2) = \left(\frac{\sigma^2}{n-1} \right)^2 \cdot \mathrm{Var}\left(\frac{(n-1) \cdot \mathrm{S}^2}{\sigma^2} \right) = \left(\frac{\sigma^2}{n-1} \right)^2 \cdot 2(n-1) = \frac{2\sigma^4}{n-1}.$$

$$\mathrm{MSE}(c\,\hat{\theta}) = \mathrm{E}\left[(c\,\hat{\theta}-\theta)^2\right] = c^2\,\mathrm{E}(\hat{\theta}^2) - 2\,c\,\mathrm{E}(\hat{\theta})\,\theta + \theta^2.$$

$$c_{\min} = \frac{\mathrm{E}(\hat{\theta})\cdot\theta}{\mathrm{E}(\hat{\theta}^2)}.$$

For $\hat{\theta} = \mathrm{S}^2$,

$$c_{\min} = \frac{\mathrm{E}(\mathrm{S}^2)\cdot\sigma^2}{\mathrm{Var}(\mathrm{S}^2) + \left[\mathrm{E}(\mathrm{S}^2)\right]^2} = \frac{\sigma^4}{\frac{2\sigma^4}{n-1}+\sigma^4} = \frac{n-1}{n+1}.$$

$\frac{n-1}{n+1}\cdot\mathrm{S}^2 = \frac{1}{n+1}\cdot\sum_{i=1}^{n}\left(\mathrm{X}_i-\overline{\mathrm{X}}\right)^2$ would minimize the Mean Squared Error.

OR

$$\mathrm{E}(c\cdot\mathrm{S}^2) = c\cdot\mathrm{E}(\mathrm{S}^2) = c\cdot\sigma^2.$$

$$\mathrm{Var}(c\cdot\mathrm{S}^2) = c^2\cdot\mathrm{Var}(\mathrm{S}^2) = c^2\cdot\frac{2\sigma^4}{n-1} = \frac{2c^2\sigma^4}{n-1}.$$

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{E}\left[(\hat{\theta}-\theta)^2\right] = \left(\mathrm{E}(\hat{\theta})-\theta\right)^2 + \mathrm{Var}(\hat{\theta}) = \left(\mathrm{bias}(\hat{\theta})\right)^2 + \mathrm{Var}(\hat{\theta}).$$

$$\mathrm{MSE}(c\cdot\mathrm{S}^2) = \left(\mathrm{E}(c\cdot\mathrm{S}^2)-\sigma^2\right)^2 + \mathrm{Var}(c\cdot\mathrm{S}^2)$$

$$= (c-1)^2\cdot\sigma^4 + \frac{2c^2\sigma^4}{n-1} = \left[\frac{n+1}{n-1}c^2 - 2c + 1\right]\cdot\sigma^4.$$

$$\frac{d}{dc}\mathrm{MSE}(c\cdot\mathrm{S}^2) = 2\cdot\left[\frac{n+1}{n-1}c - 1\right]\cdot\sigma^4 = 0. \qquad \Rightarrow \qquad c = \frac{n-1}{n+1}.$$

$\frac{n-1}{n+1}\cdot\mathrm{S}^2 = \frac{1}{n+1}\cdot\sum_{i=1}^{n}\left(\mathrm{X}_i-\overline{\mathrm{X}}\right)^2$ would minimize the Mean Squared Error.

$$\left(\frac{d^2}{dc^2}\mathrm{MSE}(c\cdot\mathrm{S}^2) = 2\cdot\left[\frac{n+1}{n-1}\right]\cdot\sigma^4 > 0 \quad \Rightarrow \quad \text{min.}\right)$$

5. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a distribution with probability density function

$$f_X(x) = f_X(x;\theta) = \frac{1+\theta x}{2}, \qquad -1 < x < 1, \qquad -1 < \theta < 1.$$

a) Obtain the method of moments estimator of θ, $\tilde{\theta}$.

$$\mu = E(X) = \int_{-1}^{1} x \cdot \frac{1+\theta x}{2}\, dx = \left(\frac{x^2}{4} + \frac{\theta x^3}{6} \right)\Bigg|_{-1}^{1} = \frac{\theta}{3}.$$

$$\overline{X} = \frac{\tilde{\theta}}{3} \qquad \Rightarrow \qquad \tilde{\theta} = 3\,\overline{X}.$$

b) Is $\tilde{\theta}$ an unbiased estimator for θ? *Justify your answer.*

$$E(\tilde{\theta}) = E(3\,\overline{X}) = 3\,E(\overline{X}) = 3\,\mu = 3\,\frac{\theta}{3} = \theta.$$

$\Rightarrow$ $\tilde{\theta}$ an unbiased estimator for θ.

c) Find $\operatorname{Var}(\tilde{\theta})$.

$$E(X^2) = \int_{-1}^{1} x^2 \cdot \frac{1+\theta x}{2}\, dx = \left(\frac{x^3}{6} + \frac{\theta x^4}{8} \right)\Bigg|_{-1}^{1} = \frac{1}{3}.$$

$$\sigma^2 = \operatorname{Var}(X) = \frac{1}{3} - \left(\frac{\theta}{3}\right)^2 = \frac{3-\theta^2}{9}.$$

$$\operatorname{Var}(\tilde{\theta}) = 9\,\operatorname{Var}(\overline{X}) = 9 \cdot \frac{\sigma^2}{n} = \frac{3-\theta^2}{n}. \qquad \Rightarrow \qquad \operatorname{MSE}(\tilde{\theta}) = \frac{3-\theta^2}{n}.$$

6. Let X_1, X_2 be a random sample of size $n = 2$ from a distribution with probability density function

$$f_X(x) = f_X(x;\theta) = \frac{1+\theta x}{2}, \qquad -1 < x < 1, \qquad\qquad -1 < \theta < 1.$$

Find the maximum likelihood estimator $\hat{\theta}$ of θ.

$$L(\theta) = \frac{1+\theta x_1}{2} \cdot \frac{1+\theta x_2}{2} = \frac{1+\theta(x_1+x_2)+\theta^2 x_1 x_2}{4}$$

$L(\theta)$ is a parabola with vertex at $\dfrac{-b}{2a} = \dfrac{-(x_1+x_2)}{2x_1x_2}$.

Case 1: $\quad a = x_1 x_2 > 0. \qquad$ Parabola has its "antlers" up.

$\Rightarrow \qquad$ The vertex is the minimum.

Subcase 1: $\quad x_1 > 0, \;\; x_2 > 0. \qquad$ Vertex $= -\dfrac{x_1+x_2}{2x_1x_2} < 0.$

Maximum of $L(\theta)$ on $-1 < \theta < 1$ is at $\hat{\theta} = 1.$

Subcase 2: $\quad x_1 < 0, \;\; x_2 < 0. \qquad$ Vertex $= -\dfrac{x_1+x_2}{2x_1x_2} > 0.$

Maximum of $L(\theta)$ on $-1 < \theta < 1$ is at $\hat{\theta} = -1.$

Case 2: $\quad a = x_1 x_2 < 0. \qquad$ Parabola has its "antlers" down.

$\Rightarrow \qquad$ The vertex is the maximum.

Vertex is at $-\dfrac{x_1+x_2}{2x_1x_2}$.

Subcase 1: $\quad -\dfrac{x_1 + x_2}{2x_1 x_2} > 1$. That is, $x_2 > -\dfrac{x_1}{2x_1 + 1}$.

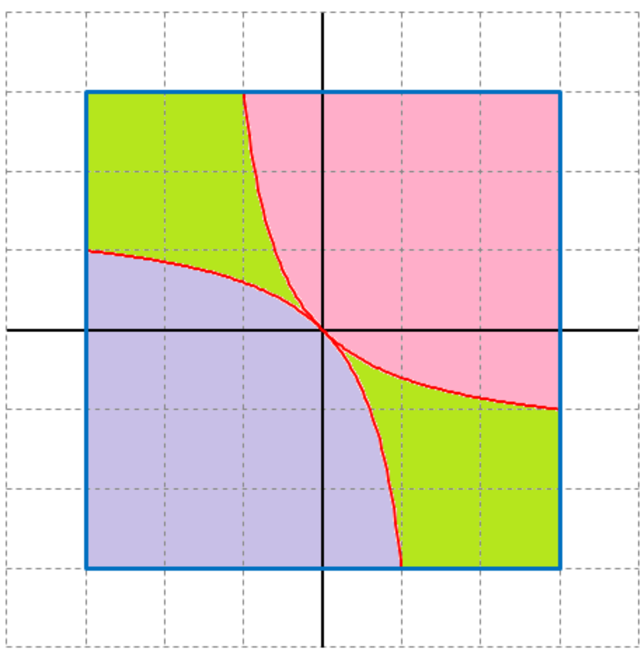
Maximum of $L(\theta)$ on $-1 < \theta < 1$ is at $\hat{\theta} = 1$.

Subcase 2: $\quad -\dfrac{x_1 + x_2}{2x_1 x_2} < -1$. That is, $x_2 < \dfrac{x_1}{2x_1 - 1}$.

Maximum of $L(\theta)$ on $-1 < \theta < 1$ is at $\hat{\theta} = -1$.

Subcase 3: $\quad -1 < -\dfrac{x_1 + x_2}{2x_1 x_2} < 1$.

Maximum of $L(\theta)$ on $-1 < \theta < 1$ is at $\hat{\theta} = -\dfrac{X_1 + X_2}{2X_1 X_2}$.



$-1 < x_1 < 1$,

$-1 < x_2 < 1$,

Pink $\quad \hat{\theta} = 1$.

Lavender $\quad \hat{\theta} = -1$.

Green $\quad \hat{\theta} = -\dfrac{X_1 + X_2}{2X_1 X_2}$.

Now imagine $n = 20$. Then $L(\theta)$ is a polynomial of power 20. Then $L'(\theta)$ is a polynomial of power 19. Solving $L'(\theta) = 0$ is no longer an option. We would have to use numerical methods to maximize the likelihood function for given $x_1, x_2, \dots, x_{20}$.