$1-\delta$ is the nominal coverage,

and $1-\delta \leq \liminf_n (1-\delta_n)$.

4) In a simulation, let Y_{f_i} be the ith random variable and PI_i the ith PI. $P(Y_{f_i} \in PI_i) \sim bin(k, 1-\delta_n)$ $\approx bin(k, 1-\delta)$ where k is the number of runs in the simulation.

often want 90%, 95% and 99% PIs.

a) $k = 5000$ simulated coverage $\in [0.94, 0.96]$ suggests actual coverage $1-\delta_n$ is close to the nominal coverage $1-\delta = 0.95$.

b) $k = 100$ simulated coverage $\in [0.89, 1]$ suggests actual coverage $1-\delta_n$ is close to the nominal coverage $1-\delta = 0.95$

reasoning: $X_k \sim bin(k, 1-\delta) \stackrel{D}{=} \sum_{i=1}^{k} Y_i$

$Y_i \sim bin(1, 1-\delta)$, $V(Y_i) = (1-\delta)\delta$

(Let $0.9 \leq 1-\delta \leq 0.95$, but take $\delta = .05$.)

$\frac{X_n}{n} = \bar{Y} = \hat{p}$ and want

$$\hat{p} \pm z_{1-\alpha}\sqrt{\frac{(1-\delta)\delta}{K}} = \hat{p} \pm .06 \quad \text{for } K=100$$

$\underbrace{\quad}_{\approx .95 \text{ for good } 95\% \text{ PI}}$

$K=100$ want $z_{1-\alpha}\, 0.0218 = .06$

or $z_{1-\alpha} = \frac{.06}{.0218} = 2.752$.

$$K=5000 \quad \hat{p} \pm z_{1-\alpha}\sqrt{\frac{(1-\delta)\delta}{K}} = \hat{p} \pm 0.01$$

$\underbrace{\quad}_{\approx .95 \text{ for good } 95\% \text{ PI}}$

or $z_{1-\alpha}\, .00308 = .01$

$$z_{1-\alpha} = \frac{.01}{.00308} = 3.244$$

That is pick $z_{1-\alpha}$ between 2 and 4

so that $z_{1-\alpha}\sqrt{\frac{(1-\delta)\delta}{K}}$ is a simple percentage.

5) Rule of thumb; for $K=100$,

target sample 95% PIs

look at PIs with coverage ≥ 0.89.

The PI with the shortest average length is best.

For $k = 5000$, look at PIs with coverage ≥ 0.94

The PI with shortest length is best.

Note: Can replace PIs with CIs, prediction regions, and confidence regions.

---

nominal 95% PI		a)	b)	c)	$k=100$
ex)	cov	.900	.86	.96	
	len	.400	.378	.383	
			cov < 0.89	best	

6) For percentile based PIs

$$[L_n, U_n] = \left[\hat{Y}_{\delta_1}, \hat{Y}_{\delta_2}\right] \text{ with } \delta_2 - \delta_1 \geq 1-\delta$$

give large sample $100(1-\delta)\%$ PIs.

see 7) and 9).

Need to use closed intervals, not open intervals.

ex] point mass at 0: $P(Y_f = 0) = 1$.

$[0,0]$ is a 100% PI and a large sample $100(1-\delta)\%$ PI for $0 < \delta < 1$.

$\underbrace{(0,0) = \emptyset}_{\text{empty set}}$ is a 0% PI.

7] Let $Y_1, \ldots, Y_n, Y_f$ be iid. Let $Y_{(1)} \le \cdots \le Y_{(n)}$ be the order statistics of the training data. Let $k_1 = \lceil \frac{n\delta}{2} \rceil$ and $k_2 = \lceil n(1 - \frac{\delta}{2}) \rceil$ where $0 < \delta < 1$. The large sample $100(1-\delta)\%$ percentile PI for Y_f is

$$[Y_{(k_1)}, Y_{(k_2)}].$$

8] Let the shortest closed interval containing at least c of the $Y_1, \ldots, Y_n$ be

Shorth$(c) = [Y_{(s)}, Y_{(s+c-1)}]$.

For $Y_1, \ldots, Y_n$ iid and $k_n = \lceil n(1-\delta) \rceil$

Frey showed that the Shorth(k_n) PI

has maximum undercoverage $\approx 1.12 \sqrt{\frac{\delta}{n}}$

for large ns. The maximum undercoverage

occurs for Uniform(θ_1, θ_2) dists

(where the pop shorth is not unique).

9)* Let $Y_1, \ldots, Y_n, Y_f$ be iid. The

large sample $100(1-\delta)\%$ Shorth(c) PI

for Y_f is $[Y_{(s)}, Y_{(s+c-1)}]$ where

$$c = \min\left(n, \left\lceil n \left[1 - \delta + \underbrace{1.12\sqrt{\frac{\delta}{n}}}_{\text{correction factor}}\right] \right\rceil\right).$$

Idea: want 90% PI but max undercov = 5%

so use a 95% PI for the given n.

10) Let $Y_1, \ldots, Y_n, Y_f$ be iid.

Let $W_{(1)}, \ldots, W_{(n)}$ be the order stats of $W_i = Y_i^2$. Let $k_n = \lceil n(1-\delta) \rceil$.

Let $\hat{L}_n = -\hat{U}_n$ and $\hat{U}_n = \sqrt{W_{(k_n)}}$. Then

$[\hat{L}_n, \hat{U}_n]$ is a large sample $100(1-\delta)\%$ PI

for Y_f.

Note: a) $P(0 \le W_f < \hat{U}_n^2) = P(-\sqrt{\hat{U}_n^2} \le Y_f \le \sqrt{\hat{U}_n^2})$

$= P[\hat{L}_n \le Y_f \le \hat{U}_n]$ is eventually

bounded below by $1-\delta$.

b) If $Y_f \ge 0$, use $[0, \hat{U}_n]$ as the P.I.

11) By Chebyshev's ineq. for $k > 1$,

$P(\mu - k\sigma \le Y \le \mu + k\sigma) \ge 1 - \frac{1}{k^2}$.

Then $k = 5$ gives $\ge 96\%$ asymptotic coverage.

12) Let $\hat{\mu} = \bar{Y}$ and $S = \sqrt{S^2}$. Let $Y_1, \ldots, Y_n, Y_f$

be iid with $E(Y) = \mu$, $SD(Y) = \sqrt{V(Y)} = \sigma$.

Let $1 - \frac{1}{k^2} \ge 1-\delta$. Let $\mu \pm k\sigma$ be continuity

points of $F(y)$. Then $[\hat{L}_n, \hat{U}_n] = [\hat{\mu} - k\hat{\sigma}, \hat{\mu} + k\hat{\sigma}]$

is a large sample $100(1-\delta)\%$ Chebyshev PI.

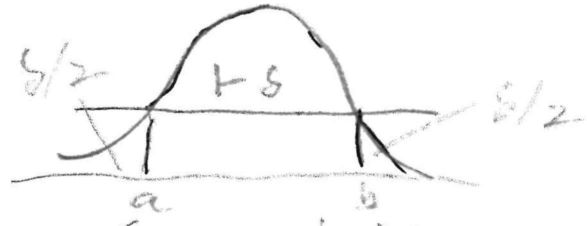
Note: This PI is usually too long if $k=5$.

$k=1.96$ is often used for a 95% PI if the data are iid $N(\mu,\sigma^2)$. This PI is usually too short for a 95% PI if the data are not normal. (MLR and Time series)

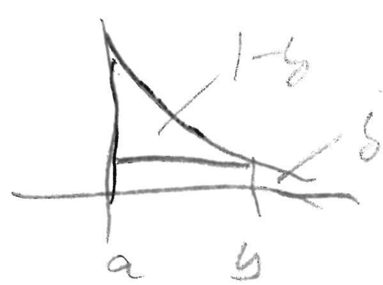
End Exam 2 material

13) The $100(1-\delta)\%$ highest density region is a union of k disjoint intervals such that the mass within the intervals $\geq 1-\delta$ and the sum of the k interval lengths is as small as possible. For a pdf, move a horizontal line down from the top of the pdf. The line will intersect the pdf or boundaries of the pdf at $[a_1, b_1], \ldots, [a_k, b_k]$ where $k \geq 1$. Stop moving the line when the areas under the pdf corresponding to the intervals $= 1-\delta$.

ex)

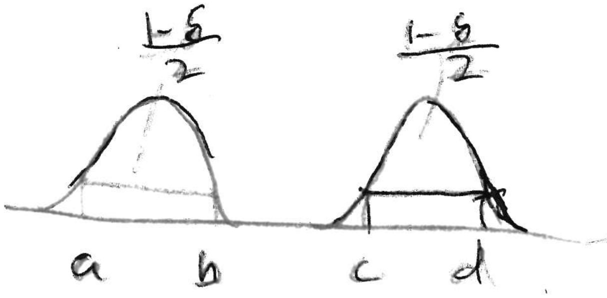


symmetric



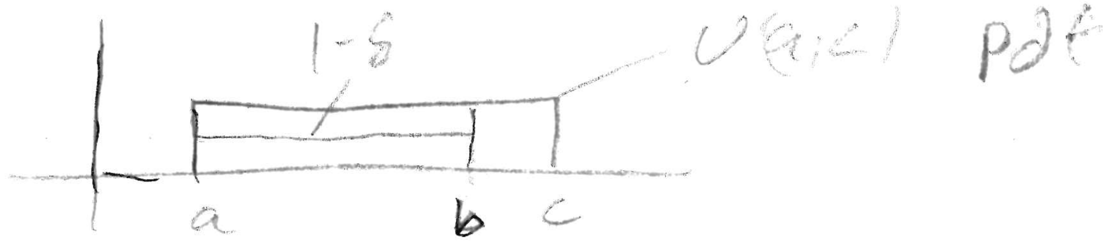
The two intervals [a,b] are also the pop shorth.

ex)



$[a,b] \cup [c,d]$ is the highest density region

ex)



[a,b] is a pop shorth and a highest density region

14) Let the training data be $\underline{x}_1, \ldots, \underline{x}_n$ and test data be $\underline{x}_f$.

A large sample $100(1-\delta)\%$ <u>prediction region</u> is a set A_n such that

$P(\underline{x}_f \in A_n)$ is eventually bounded below by $1-\delta$ as $n \to \infty$. A prediction region is asymptotically optimal if its volume converges to that of the highest density region (the minimum volume covering region) of the dist of $\underline{x}_f$.

15) Let the ith Mahalanobis distance $D_i = \sqrt{D_i^2}$ where

$$D_i^2 = D_i^2(T, C) = (\underline{x}_i - T)^T C^{-1} (\underline{x}_i - T).$$

16) The sample mean $\underline{\bar{x}} = \frac{1}{n}\sum_{i=1}^{n} \underline{x}_i$ and the sample covariance matrix

$$S = \frac{1}{n-1}\sum_{i=1}^{n} (\underline{x}_i - \underline{\bar{x}})(\underline{x}_i - \underline{\bar{x}})^T.$$

17) Let $\underbrace{\underline{x}_1, \ldots, \underline{x}_n}_{\text{training}}, \underbrace{\underline{x}_f}_{\text{test}}$ be iid,

a) If $\underline{x}_i \overset{iid}{\sim} N_p(\underline{\mu}, \Sigma)$, the large sample $100(1-\delta)\%$ MVN prediction region for $\underline{x}_f$ is

$$\{ \underline{z} : D^2_{\underline{z}}(\overline{x}, S) \le \chi^2_{p,1-\delta} \}.$$

b) The large sample $100(1-\delta)\%$ nonparametric prediction region for $\underline{x}_f$ is

$$\{ \underline{z} : D^2_{\underline{z}}(\overline{x}, S) \le D^2_{(U_n)} \}.$$

eg 90%, 80%, 50%

18) Let
$$q_n = \begin{cases} \min\left(1-\delta+0.05,\ 1-\delta+\frac{p}{n}\right) & \delta > 0.1 \\ \min\left(1-\frac{\delta}{2},\ 1-\delta+10\delta\frac{p}{n}\right) & \delta \le 0.1 \end{cases}$$

eg 95%, 99%

If $1-\delta < 0.999$ and $q_n < 1-\delta+0.001$, set $q_n = (1-\delta)$. Let $D^2_{(U_n)}$ be the $100\, q_n$th sample percentile of $D_i^2 = D^2_{\underline{x}_i}$ for $i = 1, \ldots, n$.

So $D^2_{(U_n)} \ge D^2_{(\lceil n(1-\delta)\rceil)}$ and $D^2_{(U_n)} \downarrow D^2_{(\lceil n(1-\delta)\rceil)}$ quickly as $n \to \infty$.

19) The nonparametric prediction region starts to give good coverage for $n \geq 10p$ and good volume for $n \geq 50p$. Like the CLT, there are always distr's with nonsingular $cov(\underline{x}) = \Sigma_{\underline{x}}$ such that large n is needed.

20) If (T, C) is a consistent estimator of $(\underline{\mu}, d\,\Sigma)$ for some constant $d > 0$, Σ a nonsingular, then $D^2_{\underline{x}}(T, C) = D^2(T, C)$ is a consistent estimator of $d^{-1} D^2_{\underline{x}}(\underline{\mu}, \Sigma)$. Hence the sample percentiles of $D_i^2 = D^2_{\underline{x}_i}$ estimate the pop percentiles of $d^{-1} D^2(\underline{\mu}, \Sigma)$ at continuity points $D^2_{1-\delta}$ of the cdf of $D^2 = D^2_{\underline{x}}(\underline{\mu}, \Sigma)$.

proof) a) $D^2(T, C) = (\underline{x} - T)^T C^{-1} (\underline{x} - T) =$

4 $(x-\mu)^T$ terms
2 $\mu - T)^T C^{-1}$ terms
can get 5 terms

$$(\underline{x} - \underline{\mu} + \underline{\mu} - T)^T [C^{-1} - d^{-1}\Sigma^{-1} + d^{-1}\Sigma^{-1}] (\underline{x} - \underline{\mu} + \underline{\mu} - T)$$

$$= (\underline{x} - \underline{\mu})^T (d^{-1}\Sigma^{-1})(\underline{x} - \underline{\mu}) + \underbrace{(\underline{x} - \underline{\mu})^T}_{O_P(1)} \underbrace{(C^{-1} - d^{-1}\Sigma^{-1})}_{o_P(1)} \underbrace{(\underline{x} - \underline{\mu})}_{O_P(1)}$$

$$+ \underbrace{(\mu - T)^T}_{o_p(1)} \underbrace{C^{-1}}_{O_p(1)} \underbrace{(\mu - T)}_{o_p(1)} + \underbrace{(\mu - T)^T}_{o_p(1)} \underbrace{C^{-1}}_{O_p(1)} \underbrace{(x - \mu)}_{O_p(1)} + \underbrace{(x-\mu)^T}_{O_p(1)} \underbrace{(C^{-1} - d^{-1}\Sigma^{-1})}_{o_p(1)} \underbrace{(\mu - T)}_{o_p(1)}$$

$$+ \underbrace{(x - \mu)^T}_{O_p(1)} \underbrace{(C^{-1} - d^{-1}\Sigma^{-1})}_{o_p(1)} \underbrace{(\mu - T)}_{o_p(1)} = d^{-1} (x - \mu)^T \Sigma^{-1} (x - \mu) + o_p(1)$$

$$= d^{-1} D^2_{x}(\mu, \Sigma) + o_p(1).$$

b) Since C and Σ are nonsingular,

$D^2_{x}(T, C)$ is a continuous function of (T, C),

$\therefore D^2_{x}(T, C) \xrightarrow{P} d^{-1} D^2_{x}(\mu, \Sigma)$.

21) a) $(\bar{x}, S) \xrightarrow{P} (\mu, \Sigma) = (E(x), \operatorname{cov}(x))$ with $d = 1$.

b) If $x_i \sim N_p(\mu, \Sigma)$, then $D^2_{x_i}(\bar{x}, S) \xrightarrow{D} \chi^2_p$.

c) If $\Sigma > 0$, if $\sqrt{n}(T - \mu) \xrightarrow{D} N_p(0, \Sigma)$, and if

$C > 0$ is a consistent estimator of Σ, then

$$n(T - \mu)^T C^{-1} (T - \mu) \xrightarrow{D} \chi^2_p.$$

0ch5 Confidence Regions and the Bootstrap

1) p220-221 Let $Y = (Y_1, \ldots, Y_n)^T$ be the data. $[L_n, U_n] =$