2020-02-24

# Prediction Intervals $\quad X \sim N(\mu, \sigma^2)$

Consider $\bar{X}$ a pt estimator for population mean $\mu$.
We can use $\bar{X}$ to predict new values for $X$ in addition to estimating $\mu$.

① Known $\mu$, known $\sigma^2$

We seek $\ell, u$ s.t. $P(\ell < X < u) = 1-\alpha$.
Since $X \sim N(\mu, \sigma^2)$, $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$.
Pick $z_{\alpha/2}$ as before. Then

$$P(\mu - z_{\alpha/2}\sigma < X < \mu + z_{\alpha/2}\sigma) = 1-\alpha.$$

② Unknown $\mu$, known $\sigma^2$.

Write $X_p$ for the single predicted future value.

$X_p \sim N(\mu, \sigma^2)$ and $\bar{X} \sim N(\mu, \sigma^2/n)$. It follows that

$X_p - \bar{X} \sim N(0, \sigma^2[1+\frac{1}{n}])$ and $\frac{X_p - \bar{X}}{\sigma\sqrt{1+\frac{1}{n}}} \sim N(0,1)$.

Exercise: Write out the prediction interval.

③ Unknown $\mu$, unknown $\sigma^2$ $\quad \frac{X_p - \bar{X}}{s\sqrt{1+\frac{1}{n}}} \sim t(n-1)$

Combining ② with our derivation for the T-distr. yields

# Relative Efficiency

We can measure which of two estimators $\hat{\vartheta}_1, \hat{\vartheta}_2$ is better through the ratio

$$\frac{MSE(\hat{\vartheta}_2)}{MSE(\hat{\vartheta}_1)}.$$ (large values favor the first; small favors 2nd)

If both are unbiased, this ratio reduces to $\frac{\mathbb{V}\hat{\vartheta}_2}{\mathbb{V}\hat{\vartheta}_1}$.
We define the relative efficiency of $\hat{\vartheta}_1$ relative to $\hat{\vartheta}_2$ as

$$\mathrm{eff}(\hat{\vartheta}_1, \hat{\vartheta}_2) = \frac{\mathbb{V}\hat{\vartheta}_2}{\mathbb{V}\hat{\vartheta}_1}$$

**Example** Consider $\Sigma(X_i-\bar{X})^2$. Since $\frac{\Sigma(X_i-\bar{X})^2}{\sigma^2} \sim \chi^2(n-1)$, it follows that

$$\frac{1}{\sigma^4}\mathbb{V}\left[\Sigma(X_i-\bar{X})^2\right] = \mathbb{V}\,\chi^2(n-1) = 2\cdot(n-1)$$

Hence $$\mathbb{V}S^2 = \mathbb{V}\left[\frac{1}{n-1}\Sigma(X_i-\bar{X})^2\right] = \frac{1}{(n-1)^2}\mathbb{V}\left[\Sigma(X_i-\bar{X})^2\right] = \frac{2(n-1)\sigma^4}{(n-1)^2} =$$

$$= \frac{2\sigma^4}{n-1}.$$

Example cont.

If the normal sample has even number of observations $n=2k$, then

$$\hat{\sigma}^2 = \frac{1}{2k}\sum_i (X_{2i}-X_{2i-1})^2 \text{ estimates } \sigma^2.$$

Claim $\hat{\sigma}^2$ is unbiased

$$\mathbb{E}\hat{\sigma}^2 = \frac{1}{2k}\sum \mathbb{E}(X_{2i}-X_{2i-1})^2$$

$X_{2i}-X_{2i-1} \sim N(0, 2\sigma^2)$ and the summands are iid. Hence

$$\mathbb{E}[(X_{2i}-X_{2i-1})^2] = \mathbb{V} + \big(\underbrace{\mathbb{E}[X_{2i}-X_{2i-1}]}_{=0}\big)^2 =$$

Hence

$$\mathbb{E}\left[\sum_{i=1}^{k}(X_{2i}-X_{2i-1})^2\right] = 2k\cdot\sigma^2 + \text{covariances} = 2k\sigma^2. \qquad \blacksquare$$

Next, to figure out $\mathbb{V}\hat{\sigma}^2$, we first notice that

$$\frac{X_{2i}-X_{2i-1}}{\sqrt{2}\,\sigma} \sim N(0,1). \quad \text{So } \sum\left(\frac{X_{2i}-X_{2i-1}}{\sigma\sqrt{2}}\right)^2 \sim \chi(k)$$

and thus $\mathbb{V}\left[\sum\left(\frac{X_{2i}-X_{2i-1}}{\sigma\sqrt{2}}\right)^2\right] = 2\cdot k$

$$\text{So } 2k = \mathbb{V}\left[\frac{1}{2\sigma^2}\sum(X_{2i}-X_{2i-1})^2\right] = \frac{1}{4\sigma^4}\mathbb{V}\left[2k\hat{\sigma}^2\right] = \frac{4k^2}{4\sigma^4}\mathbb{V}\hat{\sigma}^2.$$

It follows that $\mathbb{V}\hat{\sigma}^2 = \dfrac{2\sigma^4}{k} = \dfrac{4\sigma^4}{n}.$

Example cont. To summarize:

$$\mathbb{V} S^2 = \frac{2\sigma^4}{n-1} \qquad \mathbb{V} \hat{\sigma}^2 = \frac{4\sigma^4}{n}$$

So $\text{eff}(S^2, \hat{\sigma}^2) = \dfrac{4\sigma^4/n}{2\sigma^4/(n-1)} = \dfrac{2(n-1)}{n} > 1$ for $n \geq 2$.

$\hat{\sigma}^2$ only works for even sample sizes, and has worse relative efficiency — but can be used for online estimation!

---

Consistency

We should do better if we have more data, right?

How are $\hat{\vartheta}_1 = \bar{X}$ and $\hat{\vartheta}_2 = \dfrac{X_1 + X_n}{2}$ different?

For these questions we are no longer dealing with a single estimator, but with a family of estimators:

$$\left\{ \hat{\vartheta}_n(X_1, \dots, X_n) \right\}$$

The way we will take goes through a new type of convergence: convergence in probability.

**Def** A sequence $X_1, \dots$ converges in probability to a value $x$ if for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|X_n - x| \le \varepsilon) = 1 \quad \text{or equivalently}$$

$$\lim_{n\to\infty} \mathbb{P}(|X_n - x| > \varepsilon) = 0.$$

**Def** A family $\{\hat{\vartheta}_n\}$ of estimators is consistent estimator of $\vartheta$ if $\hat{\vartheta}_n \xrightarrow{P} \vartheta$.

For unbiased estimators, it is enough to inspect the variance to show consistency:

**Theorem** A family $\{\hat{\vartheta}_n\}$ of unbiased estimators is consistent if

$$\lim_{n\to\infty} \mathbb{V}\hat{\vartheta}_n = 0.$$

**Proof** Chebyshev's theorem applied to $\hat{\vartheta}_n$ for a fixed $n$ tells us

$$\mathbb{P}(|\hat{\vartheta}_n - \vartheta| > k \cdot \sigma_{\hat{\vartheta}_n}) \le \frac{1}{k^2} \qquad \text{where } \sigma_{\hat{\vartheta}_n} = \sqrt{\mathbb{V}\hat{\vartheta}_n}.$$

For any $\varepsilon > 0$ pick $k = \frac{\varepsilon}{\sigma_{\hat{\vartheta}_n}}$. Then

$$0 \le \mathbb{P}(|\hat{\vartheta}_n - \vartheta| > \varepsilon) = \mathbb{P}\left(|\hat{\vartheta}_n - \vartheta| > \left(\frac{\varepsilon}{\sigma_{\hat{\vartheta}_n}}\right)\sigma_{\hat{\vartheta}_n}\right) \le \frac{1}{(\varepsilon/\sigma_{\hat{\vartheta}_n})^2} = \frac{\mathbb{V}\hat{\vartheta}_n}{\varepsilon^2}.$$

Since $\mathbb{P}(|\hat{\vartheta}_n - \vartheta| > \varepsilon)$ is bounded below by $0$ and above by $\frac{\mathbb{V}\hat{\vartheta}_n}{\varepsilon}$, if $\mathbb{V}\hat{\vartheta}_n \to 0$ then so does $\mathbb{P}(|\hat{\vartheta}_n - \vartheta| > \varepsilon)$. ∎

In direct analogy to how we prove limit laws with $\varepsilon$-$\delta$-proofs, we can prove the same limit laws for convergence in probability.

<u>Note</u> I suspect but haven't proved that it is enough with $Bias(\hat{\vartheta}_n) \to 0$ for the theorem to hold.

Convergence in probability is the motivation for statements like
"For large $n$, we can approximate $\sigma^2$ by $S^2$ well enough to treat the T-score as a Z-score":

<u>Theorem</u> If $U_n$ converges to $N(0,1)$ in distribution and $W_n \xrightarrow{P} 1$, then $U_n/W_n$ converges to $N(0,1)$

<u>Exercise</u> 9.26 $Y_1, \dots, Y_n \sim Uniform(0, \vartheta)$

We have seen that
$$F_{(n)}(y) = CDF_{Y_{(n)}}(y) = \begin{cases} 0 & y < 0 \\ (y/\vartheta)^n & 0 \le y \le \vartheta \\ 1 & y > \vartheta \end{cases}$$

- $\mathbb{P}(|Y_{(n)} - \vartheta| \le \varepsilon) = \mathbb{P}(\vartheta - \varepsilon \le Y_{(n)} \le \vartheta + \varepsilon)$
- If $\varepsilon > \vartheta$ then $\vartheta - \varepsilon < 0$ and $\vartheta + \varepsilon > \vartheta$, so $\mathbb{P} = 1$.
- For $\varepsilon < \vartheta$, $\mathbb{P} = \underbrace{\mathbb{P}(Y_{(n)} \le \vartheta + \varepsilon)}_{=1} - \mathbb{P}(Y_{(n)} \le \vartheta - \varepsilon) = 1 - \left(\frac{\vartheta - \varepsilon}{\vartheta}\right)^n$
- Since $\frac{\vartheta - \varepsilon}{\vartheta} < 1$, $\lim_{n \to \infty} \left(\frac{\vartheta - \varepsilon}{\vartheta}\right)^n = 0$. Hence consistent. ∎