Chapter 6.3 Inferences based on the MLE

Use Michelson - Newcomb data set; a boxplot, then a histogram without two outliers: $x=-2$, $x=-44$; $n=66$ measurements. Accept normal location-scale model. How to conduct inferences about μ and σ^2 here? For now, we will rely on the specific features of the normal distribution to do that.

6.3.1 Standard errors, Bias, Consistency

Our goal is to estimate parameter θ. We want to know how concentrated $\hat{\theta}(s)$ is around the true value θ for different samples; thus, we need to study the sampling distribution of $\hat{\theta}(s)$. Since the true θ is not known, we have to study the sampling distribution of $\hat{\theta}(s)$ for $\forall \theta \in \Omega$

Def. MSE of an estimator $\hat{\theta}$ of θ is

$$MSE_\theta(\hat{\theta}) = E_\theta(\hat{\theta}-\theta)^2 \text{ for } \forall \theta \in \Omega$$

Th. If $E_\theta \hat{\theta}$ exists, then

$$MSE_\theta(\hat{\theta}) = Var_\theta \hat{\theta} + \left[E_\theta(\hat{\theta}) - \theta\right]^2$$

The proof is directly by definition.

Def. The bias in the estimator $\hat{\theta}$ of θ is given by $E_\theta(\hat{\theta}) - \theta$ whenever $E_\theta(\hat{\theta})$ exists. If the bias is equal to 0 for $\forall\theta$, $\hat{\theta}$ is an unbiased estimator of θ.

Ex. Location Normal model

$$L(\mu | x_1, \ldots, x_n) = e^{-\frac{n}{2\sigma_0^2}(\bar{x}-\mu)^2}$$; here $\sigma_0^2 > 0$ is known. Recall that the MLE of μ is $\hat{\mu} = \bar{x}$.

We know that $\bar{X} \sim N(\mu, \frac{\sigma_0^2}{n})$ — sampling distribution and so $\bar{X}$ is unbiased. Next, $MSE_\mu(\bar{X}) = Var_\mu(\bar{X}) = \sigma_0^2/n$ which is independent of μ. Thus, $MSE_\mu(\bar{X})$ need not be estimated in this case. The $SD = \frac{\sigma_0}{\sqrt{n}} = SD_\mu(\bar{X})$.

<u>Ex. 6.3.2</u> Bernoulli model

Let $X_1, \ldots, X_n \sim Ber(\theta)$, $\theta \in [0,1]$ and unknown

Let $\mathcal{L}(\theta | x_1, \ldots, x_n) = \theta^{n\bar{x}}(1-\theta)^{n(1-\bar{x})}$;

check that the MLE is $\bar{x}$. Then, $E_\theta(\bar{X}) = \theta$ $\forall \theta \in [0,1]$,

$\bar{X}$ is an unbiased estimator of θ. Thus,

$$MSE_\theta(\bar{X}) = V_\theta(\bar{X}) = \frac{\theta(1-\theta)}{n},$$

and the estimated $MSE_{\hat{\theta}}(\bar{X}) = \frac{\hat{\theta}(1-\hat{\theta})}{n} = \frac{\bar{x}(1-\bar{x})}{n}$;

$SD_{\hat{\theta}}(\bar{X}) = \sqrt{\frac{\bar{x}(1-\bar{x})}{n}}$ — note how different this is from the previous SD of $\bar{x}$!!

<u>Ex 6.3.3</u> Estimating proportion of households in a district that will participate in a recycling program.

Take sample of $n = 1{,}000$ where $N \simeq 1.5$ mln households.

The only two possible answers are yes or no;

since $n \ll N$, assume Bernoulli(θ) model, $\theta \in [0,1]$.

Answers: 790 - yes, 210 - no.

$$\Rightarrow \hat{\theta} = \bar{x} = \frac{790}{1000} = 0.79; \quad SD_{\hat{\theta}}(\bar{x}) = \sqrt{\frac{\bar{x}(1-\bar{x})}{1000}} = 0.01288$$

Right now, we aren't sure how to interpret $SD_{\hat{\theta}}(\bar{x})$ — looking forward to CIs.

Ex 6.3.4 Location-scale normal model.

Again, $(X_1, \ldots, X_n) \sim N(\mu, \sigma^2)$; both μ and σ^2 are unknown. The parameter is $\theta = (\mu, \sigma^2) \in \Omega = R^1 \times (0, +\infty)$

Need to estimate $\mu = \psi(\mu, \sigma^2)$

$$\Rightarrow L(\mu, \sigma^2 | x_1, \ldots, x_n) = (2\pi\sigma^2)^{-n/2} e^{-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2} e^{-\frac{n-1}{2\sigma^2} s^2}$$

We also showed that the MLE of θ is

$\left(\bar{x}, \frac{n-1}{n} s^2\right)$. Furthermore, $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ independent of

$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$. The plug-in MLE of μ is $\hat{\mu} = \bar{x}$,

$\hat{\mu}$ is unbiased, $E_\theta(\bar{X}) = \frac{\sigma^2}{n}$. We must estimate

$$MSE_\theta(\bar{X}) = \frac{\sigma^2}{n} \text{ with } \frac{\frac{n-1}{n} s^2}{n} = \frac{n-1}{n^2} s^2 \approx \frac{s^2}{n}$$

This is a useful estimator of $\frac{\sigma^2}{n}$ because $E_\theta(S^2) = \sigma^2$;

$\frac{s}{\sqrt{n}}$ - standard error of $\bar{x}$.

Ex. 6.3.5. Application of the location-scale normal model.

Recall $n = 30$ heights (in) of students. We found $\bar{x} = 64.577$

and $s = 2.379$. $\Rightarrow SD(\bar{x}) = \frac{s}{\sqrt{n}} = 0.43434\ldots$

Consistency of the estimators.

Def. A sequence of estimates $T_1, T_2, \ldots$ etc. is consistent (in probability) for $\psi(\theta)$ if $T_n \xrightarrow{P_\theta} \psi(\theta)$ as $n \to \infty$ for $\forall \theta \in \Omega$.

A sequence of estimates $T_1, T_2, \ldots$ is consistent almost surely for $\psi(\theta)$ if $T_n \xrightarrow{a.s.} \psi(\theta)$ as $n \to \infty$ for $\forall \theta \in \Omega$.

Now, consider $(X_1, \ldots, X_n) \sim \{f_\theta : \theta \in \Omega\}$ & let

$T_n = n^{-1} \sum_{i=1}^{n} X_i$ be the nth sample average as an estimator of

$\psi(\theta) = E_\theta X$. By WLLN & SLLN we have what we need!

6.3.2 Confidence intervals.

The interval is $C(s) = (l(s), u(s))$ based on the data s. How do we specify its endpoints to ensure it contains the true value of $\psi(\theta)$ with high probability? We can specify some $\gamma \in [0, 1]$ and require that the random interval C has the confidence property.

Def. An interval $C(s) = (l(s), u(s))$ is a γ-confidence interval for $\psi(\theta)$ if $P_\theta(\psi(\theta) \in C(s)) =$
$= P_\theta(l(s) \le \psi(\theta) \le u(s)) \ge \gamma$ for $\forall \theta \in \Omega$.

γ is the confidence level of the interval.
Such an interval must be random. Usual choices of γ are $\gamma = 0.95$ or $\gamma = 0.99$

z - confidence intervals.

Ex. Location normal model.
$(x_1, , x_n) \sim N(\mu, \sigma_0^2)$, $\mu \in R^1$, $\sigma_0^2 > 0$ is known.
Clearly, if $\mu_1 \in C(x_1, , x_n)$, and $L(\mu_2 | x_1, , x_n) \ge L(\mu_1 | x_1, , x_n)$ μ_2 must also belong to $C(x_1, , x_n)$. Therefore, $C(x_1, , x_n)$ is of the form $C(x_1, , x_n) = \{\mu : L(\mu | x_1, x_n) \ge k(x_1, , x_n)\}$ for some boundary k. Then,

$$C(x_1, , x_n) = \{\mu : e^{-\frac{n}{2\sigma_0^2}(\bar{x} - \mu_0)^2} \ge k(x_1, , x_n)\} =$$

$$= \{\mu : -\frac{n}{2\sigma_0^2}(\bar{x} - \mu_0^2) \ge \ln k(x_1, , x_n)\} =$$

$$= \{\bar{x} - k^*(x_1, , x_n)\frac{\sigma_0}{\sqrt{n}}, \bar{x} + k^*(x_1, , x_n)\frac{\sigma_0}{\sqrt{n}}\} \text{ where}$$

$k^*(x_1, , x_n) = \sqrt{-2\ln k(x_1, x_n)}$. The simplest choice is to choose k^* to be a constant and to make the interval as short as possible.

Thus, $\gamma \le P_\mu(\mu \in C(x_1, \dots, x_n)) = P_\mu\left(\bar{X} - k^* \frac{\sigma_0}{\sqrt{n}} \le \mu \le \bar{X} + k^* \frac{\sigma_0}{\sqrt{n}}\right) =$
$= P_\mu\left(\left|\frac{\bar{X}-\mu}{\sigma_0/\sqrt{n}}\right| \le k^*\right) = 1 - 2(1 - \Phi(k^*))$
$\Rightarrow \Phi(k^*) = \frac{1+\gamma}{2}$, and $k^* = z_{\frac{1+\gamma}{2}}$ is the solution.
Thus, $\left[\bar{X} - z_{(1+\gamma)/2} \frac{\sigma_0}{\sqrt{n}}, \bar{X} + z_{(1+\gamma)/2} \frac{\sigma_0}{\sqrt{n}}\right]$ is an exact

γ-confidence interval for μ. E.g. for $\gamma = 0.95$, we have

$z_{0.975} = 1.96$. The quantity $z_{\frac{1+\gamma}{2}} \frac{\sigma_0}{\sqrt{n}}$ is the margin of

error.

<u>Ex</u>. Bernoulli model

Let $X_1, \dots, X_n \sim Ber(\theta)$, $\theta \in [0, 1]$ - unknown

Recall MLE of θ is $\hat{\theta} = \bar{X}$; $SD(\bar{X})$ is est. by $\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}$
The likelihood interval here takes the form
$C(x_1, \dots, x_n) = \{\theta : \theta^{n\bar{x}}(1-\theta)^{n(1-\bar{x})} \ge k(x_1, \dots, x_n)\}$. To determine these intervals, we have to find k such that
$\theta^{n\bar{x}}(1-\theta)^{n(1-\bar{x})} = k(x_1, \dots, x_n)$. Cannot be done explicitly.. so, recall that by CLT

$$\frac{\sqrt{n}(\bar{X}-\theta)}{\sqrt{\theta(1-\theta)}} \xrightarrow{D} N(0,1); \text{ also, } \frac{\sqrt{n}(\bar{X}-\theta)}{\sqrt{\bar{X}(1-\bar{X})}} \xrightarrow{D} N(0,1)$$

$$\Rightarrow \gamma = \lim_{n\to\infty} P_\theta\left(-z_{\frac{1+\gamma}{2}} \le \frac{\sqrt{n}(\bar{X}-\theta)}{\sqrt{\bar{X}(1-\bar{X})}} \le z_{\frac{1+\gamma}{2}}\right) =$$

$$= \lim_{n\to\infty} P_\theta\left(\bar{X} - z_{\frac{1+\gamma}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \le \theta \le \bar{X} + z_{\frac{1+\gamma}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right)$$

and $\left[\bar{x} - z_{\frac{1+\gamma}{2}}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}, \bar{x} + z_{\frac{1+\gamma}{2}}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\right]$ is an approx. γ-confidence

interval for θ

In practice, n may need to be quite large for this interval to behave well. This is particularly true if θ is close to 0 or 1.

t - confidence intervals.

Ex Location - scale normal model.

Now, let $(x_1, \ldots, x_n) \sim N(\mu, \sigma^2)$; both μ, σ^2 are unknown. Here, $\theta = (\mu, \sigma^2) \in \Omega = \mathbb{R}^1 \times (0, \infty)$ and $\psi(\theta) = \psi(\mu, \sigma^2) = \mu$. The likelihood f-n here depends on (μ, σ^2) and so the direct reasoning of the location normal model won't work here. We will restrict our attention to intervals of the type $C(x_1, \ldots, x_n) = \left[\bar{x} - k\frac{s}{\sqrt{n}}, \bar{x} + k\frac{s}{\sqrt{n}}\right]$ for some k.

Then, $P_{(\mu,\sigma^2)}\left(\bar{x} - k\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + k\frac{s}{\sqrt{n}}\right) = P_{(\mu,\sigma^2)}\left(-k \le \frac{\bar{x}-\mu}{s/\sqrt{n}} \le k\right) =$

$= P_{(\mu,\sigma^2)}\left(\left|\frac{\bar{x}-\mu}{s/\sqrt{n}}\right| \le k\right) = 1 - 2\left[1 - G(k; n-1)\right]$

and $G(\cdot\,; n-1)$ is the cdf of $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$. Recall

that $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ independent of $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$.

$\Rightarrow T = \left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right) \Big/ \sqrt{\frac{(n-1)S^2}{\sigma^2}} \sim t(n-1)$.

Thus, take $k = t_{\frac{1+\gamma}{2}}(n-1)$ to obtain

$\left[\bar{x} - t_{\frac{1+\gamma}{2}}(n-1), \bar{x} + t_{\frac{1+\gamma}{2}}(n-1)\frac{s}{\sqrt{n}}\right]$ — an exact γ-conf. interval for μ.

1. These intervals are in general wider than z-conf. int.
2. These CI's are not likelihood intervals for μ. We will see later that these intervals arise from another approach to inference. For now, they are just intuitively reasonable.
3. Ex. Let $n = 30$ heights (in) of students. $\Rightarrow \bar{x} = 64.517$,

$\frac{s}{\sqrt{30}} = 0.43434$. $\Rightarrow t_{0.975}(29) = 2.0452$;

$\Rightarrow \left[64.517 \pm 2.0452(0.43434)\right] = \left[63.629, 65.405\right]$

6.3.3. Testing hypotheses and P-values.

Conjecture, say $\Psi(\theta) = \Psi_0$; typically written $H_0 : \Psi(\theta) = \theta_0$ — the null hypothesis. Usually, it is associated with the treatment having no effect. ...

Next, based on the data s, we wish to assess the evidence for $\Psi(\theta) = \Psi_0$ being true. A statistical procedure that does this is called a test of significance, or a test of hypothesis. We measure how surprising it would be to observe s were $H_0 : \Psi(\theta) = \theta_0$ true. We measure P-value that indicates the degree of surprise given H_0 is true. Note that P-value $\neq P(H_0$ is true$)$; it is not the evidence that H_0 is true.

z-tests.

Ex. 6.3.9. Location - normal model

$(x_1, \dots, x_n) \sim N(\mu, \sigma_0^2)$, $\sigma_0^2 > 0$ is known.

Take $H_0 : \mu = \mu_0$, under H_0, $\bar{X} \sim N(\mu_0, \frac{\sigma_0^2}{n})$.

Thus, if $\bar{x}$ is in a region of low probability for $N(\mu_0, \frac{\sigma_0^2}{n})$, it would be an evidence that H_0 is false. This happens if $\bar{x}$ is e.g. in the right tail of the distribution. Since $Z = \frac{\bar{X} - \mu_0}{\sigma_0/\sqrt{n}} \sim N(0,1)$ under H_0, the P-value is $P_{\mu_0}\left(|\bar{X} - \mu_0| \geq |\bar{x} - \mu_0|\right) = P_{\mu_0}\left(\left|\frac{\bar{X} - \mu_0}{\sigma_0/\sqrt{n}}\right| \geq \left|\frac{\bar{x} - \mu_0}{\sigma_0/\sqrt{n}}\right|\right)$

$= 2\left[1 - \Phi\left(\left|\frac{\bar{x} - \mu_0}{\sigma_0/\sqrt{n}}\right|\right)\right]$; this is a z-test procedure.

Ex. Application of z-test. Let $N(26, 4)$ be the distr., take $n = 10$; test the hypothesis $H_0 : \mu_0 = 25$;

P-value is $2\left[1 - \Phi\left(\frac{|\bar{x} - \mu_0|}{\sigma_0/\sqrt{n}}\right)\right] = \dots\dots = 0.0078$

In R: x <- rnorm(10, 26, 4); σ_0 is known, so pval <- 2*[1 - pnorm(|xbar-25| / (2/√10))]

xbar <- mean(x)

Ex. Bernoulli model

Let $(X_1, \dots, X_n) \sim Ber(\theta)$; $H_0: \theta = \theta_0$;

If H_0 is true, $Z = \dfrac{\sqrt{n}(\bar{X}-\theta_0)}{\sqrt{\theta_0(1-\theta_0)}} \xrightarrow{D} N(0,1)$.

$\Rightarrow$ approx. P-value is $P\left(|Z| \geq \left|\dfrac{\sqrt{n}(\bar{x}-\theta_0)}{\sqrt{\theta_0(1-\theta_0)}}\right|\right) \approx$

$\approx 2\left[1-\Phi\left(\left|\dfrac{\sqrt{n}(\bar{x}-\theta_0)}{\sqrt{\theta_0(1-\theta_0)}}\right|\right)\right]$ when n is large

Toss a coin $n=100$ times, if you guessed 54 out of 100, are you a psychic? $H_0: \theta = \theta_0$ where $\theta_0 = \frac{1}{2}$.

The MLE is 0.54, P-val. $\approx 2\left[1-\Phi\left(\left|\dfrac{\sqrt{100}\,(0.54-0.5)}{\sqrt{0.5(1-0.5)}}\right|\right)\right] \approx$

≈ 0.4238 – no evidence H_0 is false!

General cutoffs are 0.05 or 0.01 but they are not necessary!!

R-command:

```
pval <- 2*[1 - pnorm(abs(√100 (xbar - 0.5)/√0.5(1-0.5)))]
xbar <- mean(x)
x <- rbinom(100, 1, 0.5)
```

Statistical vs. Practical Significance

Let the true value $\mu_1 \neq \mu_0$ but very close to it.

By SLLN, $\bar{X} \xrightarrow{a.s} \mu_1$ as $n \to \infty$, therefore

$$\left| \frac{\bar{X} - \mu_0}{\sigma_0 / \sqrt{n}} \right| \xrightarrow{a.s.} \infty.$$

This implies that $2\left[1 - \Phi\left(\left|\frac{\bar{X} - \mu_0}{\sigma_0/\sqrt{n}}\right|\right)\right] \xrightarrow{a.s.} 0$

Thus, for a sufficiently large sample size we can make P-value as close to 0 as needed.

This has nothing to do with the practical significance.

One possibility is to consider the observed absolute difference $|\bar{X} - \mu_0|$ (as an estimate of $|\mu - \mu_0|$) if this absolute difference is less than some threshold, selected in accordance with practical needs, P-value is not relevant to our decision-making.

Hypothesis assessment via confidence intervals.

e.g. data from Bernoulli model, for 95% CI $\bar{x} \pm z_{0.975}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}} = 0.54 \pm 1.96\sqrt{\frac{0.54(1-0.54)}{100}} = [0.44233, 0.63769]$

If $C(s)$ is a CI for $\psi(\theta)$, and $\psi_0 \notin C(s)$, this seems to be the evidence against $H_0 : \psi = \psi_0$, at least when γ is close to 1.

Testing through the use of P-value with a specific cutoff is equivalent to the use of a CI with a given significance.

Ex Let γ be the confidence level of a z-interval, we will show that obtaining a P-value $< 1-\gamma$ for $H_0 : \mu = \mu_0$ is equivalent to μ_0 not being in a CI for μ of level γ.

Observe that

$$1 - \gamma \leq 2\left[1 - \Phi\left(\left|\frac{\bar{X} - \mu_0}{\sigma_0/\sqrt{n}}\right|\right)\right] \text{ is } \Leftrightarrow$$

$$\Phi\left(\left|\frac{\bar{X} - \mu_0}{\sigma_0/\sqrt{n}}\right|\right) \leq \frac{1+\gamma}{2} \Leftrightarrow \left|\frac{\bar{X} - \mu_0}{\sigma_0/\sqrt{n}}\right| \leq z_{\frac{1+\gamma}{2}} \Rightarrow$$

$$\Leftrightarrow \mu_0 \in \left[\bar{X} - z_{\frac{1+\gamma}{2}}\frac{\sigma_0}{\sqrt{n}}, \bar{X} + z_{\frac{1+\gamma}{2}}\frac{\sigma_0}{\sqrt{n}}\right]$$

Thus, the γ- CI for μ comprises those values μ_0 for which the P-value for $H_0: \mu = \mu_0$ is > than $1-\gamma$. Therefore,

the P-value, based on the z-statistic, for the $H_0: \mu = \mu_0$, will be smaller then $1-\gamma$, <=> μ_0 is not in the CI for μ that we derived earlier. Ex. 6.3.10 – check the CI using R.

t-tests.

Ex Location - Scale normal model and t-tests.

Let $x_1, x_n \sim N(\mu, \sigma^2)$ and $H_0: \mu = \mu_0$.

We can prove that under H_0, $T = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} \sim t_{n-1}$.

Thus, P-val. $= P_{(\mu_0, \sigma^2)}\left(|T| \geq \left|\frac{\bar{x}-\mu_0}{s/\sqrt{n}}\right|\right) =$

$= 2\left[1 - G\left(\left|\frac{\bar{x}-\mu_0}{s/\sqrt{n}}\right|; n-1\right)\right]$ where $G(\cdot; n-1)$ is the cdf of t_{n-1}. – t-test. Again, it is advisable to look at $|\bar{x}-\mu_0|$ as an estimate of $|\mu-\mu_0|$

– can use Ex 6.3.10 data again to conduct this test. $n=10$, $N(26,4)$: x <- rnorm(10, 26, 4), $H_0: \mu = 25$

xbar <- mean(x); SD <- sqrt(var(x))

t <- $\frac{\text{xbar} - 25}{\text{SD}/\text{sqrt}(10)}$; pval <- 2*pt(-abs(t), df=9)

One-sided tests.

Sometimes, it is necessary to test $H_0: \psi(\theta) \leq \psi_0$ or $H_0: \psi(\theta) \geq \psi_0$. – one-sided tests.

Ex. $x_1, x_n \sim N(\mu, \sigma_0^2)$, $\sigma_0^2 > 0$ is known.
& $H_0: \mu \leq \mu_0$. Base the test on the z-statistic

$$Z = \frac{\bar{x}-\mu_0}{\sigma_0/\sqrt{n}} = \frac{\bar{x}-\mu+\mu-\mu_0}{\sigma_0/\sqrt{n}} = \frac{\bar{x}-\mu}{\sigma_0/\sqrt{n}} + \frac{\mu-\mu_0}{\sigma_0/\sqrt{n}} \text{ - constant}$$

Thus, $Z \sim N\left(\frac{\mu-\mu_0}{\sigma_0/\sqrt{n}}, 1\right)$;

Note that $\frac{\mu-\mu_0}{\sigma_0/\sqrt{n}} \leq 0 \iff H_0$ is true.

Thus, if H_0 is false, we tend to see values of Z in the right tail of $N(0,1)$. Compute P-value

$$P\left(Z \geq \frac{\bar{X}-\mu_0}{\sigma_0/\sqrt{n}}\right) = 1 - \Phi\left(\frac{\bar{X}-\mu_0}{\sigma_0/\sqrt{n}}\right)$$; if it is small – reject H_0. By the same reasoning,

$$\text{if } H_0: \mu \geq \mu_0, \ P\text{-val.} = P\left(Z \leq \frac{\bar{X}-\mu_0}{\sigma_0/\sqrt{n}}\right) = \Phi\left(\frac{\bar{X}-\mu_0}{\sigma_0/\sqrt{n}}\right).$$

6.3.4. Inferences for the variance

So far, we concentrated on μ in $N(\mu,\sigma^2)$ or θ in $Ber(\theta)$; often, σ^2 in $N(\mu,\sigma^2)$ is referred to as the nuisance parameter. Note that in $Ber(\theta)$, since the variance is $\theta(1-\theta)$, there are no nuisance parameters.

Sometimes, however, σ^2 is the object of study on its own. Thus, we may want to test $H_0: \sigma^2 = \sigma_0^2$. Note that S^2 is a very natural estimator of σ^2.

Ex. Let $X_1, \ldots, X_n \sim N(\mu,\sigma^2)$. The plug-in MLE of σ^2 is $\frac{n-1}{n}S^2$; S^2 is unbiased and we will use it here, however.

Recall that $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$; let $\chi^2_\alpha(\lambda)$ denote the αth quantile for $\chi^2(\lambda)$ distr., then for $\gamma \in [0,1]$

$$\gamma = P_{(\mu,\sigma^2)}\left(\chi^2_{\frac{1-\gamma}{2}}(n-1) \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{\frac{1+\gamma}{2}}(n-1)\right) =$$

$$= P_{(\mu,\sigma^2)}\left(\frac{(n-1)S^2}{\chi^2_{\frac{1+\gamma}{2}}(n-1)} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{\frac{1-\gamma}{2}}(n-1)}\right)$$ for any $(\mu,\sigma^2) \in R^1 \times (0,\infty)$.

So, $\left(\frac{(n-1)S^2}{\chi^2_{\frac{1+\gamma}{2}}(n-1)}, \frac{(n-1)S^2}{\chi^2_{\frac{1-\gamma}{2}}(n-1)}\right)$ is an exact γ-CI for σ^2.

To test $H_0: \sigma = \sigma_0$ at the $1-\gamma$ level, just check whether or not σ_0^2 is in the interval.

Again, Ex 6.3.10 data: pretend you don't know $\sigma^2 = 4$.

```
x <- rnorm(10, 26, 4)
SD <- sqrt(var(x))
V <- var(x)
0.95% CI:
LB <- 9*V / qchisq(.025, df=9)
UB <- 9*V / qchisq(0.975, df=9)
```

; $\gamma = 0.95$, so need $\chi^2_{0.975}(9)$ and $\chi^2_{0.025}(9)$,

6.3.5. Sample size calculations: CI's.

Ex 6.3.16 (i) Guaranteeing that the CI is sufficiently narrow; let $X_1, \ldots, X_n \sim N(\mu, \sigma_0^2)$ & $\sigma_0^2 > 0$ is known.

Choose n such that $z_{\frac{1+\gamma}{2}} \frac{\sigma_0}{\sqrt{n}} \le \delta$ for fixed level δ.

$\Rightarrow n \ge \sigma_0^2 \left(\frac{z_{(1+\gamma)/2}}{\delta} \right)^2$; e.g. if $\sigma_0^2 = 10$, $\gamma = 0.95$, $\delta = 0.5$, $\Rightarrow$ smallest possible n is 154.

(ii) if $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, $\sigma^2 > 0$ unknown, we can say technically that $n \ge S^2 \left(\frac{t_{(1+\gamma)/2}(n-1)}{\delta} \right)^2$ but this is impossible to compute beforehand.

Usually, we can find b s.t. $\sigma \le b$: e.g., if the data is Gaussian, we can assume that the data is within $\mu \pm 3\sigma$

$\Rightarrow$ divide the range of the data by 6 to find b) $\Rightarrow$

$\Rightarrow$ assume that $s \le b$, thus, can choose

$$n \ge b^2 \left(\frac{t_{(1+\gamma)/2}(n-1)}{\delta} \right)^2$$

Conservative choice of n is very important everywhere!!

Ex. 6.3.17. Length of a CI for proportion.
Let $(X_1, \dots, X_n) \sim Ber(\theta)$, $\theta \in [0,1]$ unknown.
Thus, we want $z_{\frac{1+\gamma}{2}} \sqrt{\frac{\bar{x}(1-\bar{x})}{n}} \le \delta$,

$$\Rightarrow n \ge \bar{x}(1-\bar{x}) \left(\frac{z_{(1+\gamma)/2}}{\delta} \right)^2$$ — unobserved $\bar{x}$ is a problem.

However, since $0 \le \bar{x} \le 1$, $0 \le \bar{x}(1-\bar{x}) \le \frac{1}{4}$, the upper bound is achieved when $\bar{x} = \frac{1}{2}$. Thus, we determine n such that $n \ge \frac{1}{4}\left(\frac{z_{(1+\gamma)/2}}{\delta} \right)^2$. e.g. $\gamma = 0.95$, $\delta = 0.1$, the smallest possible value of n is 97.

6.3.6. Sample-size calculations: power.
Need to choose n such that P-val. $\le \alpha$, with prob. at least β_0, at a specific θ_1 s.t. $\psi(\theta_1) \ne \psi_0$ while $H_0: \psi(\theta) = \psi_0$. The prob. that P-val. $\le \alpha$ for a given alternative θ_1 is the power of the test at θ_1. We will use the notation $\beta(\theta)$ — the power function. Thus, the problem is to find n s.t. $\beta(\theta_1) \ge \beta_0$ for some threshold β_0. Note that $\beta(\theta)$ in fact depends also on α, n etc. but we suppress this for simplicity.

For any test procedure, it is a good idea to examine its power function for several choices of α to see how good the test is at detecting departures. Always choose ψ_1 that represents a practically significant departure from ψ_0 and then determine n so that we reject H_0 with high probability when $\psi(\theta) = \psi_1$.

Ex. 6.3.18. Location normal model
For a 2-sided z-test, $\beta(\mu) = P_\mu\left(2\left[1 - \Phi\left(\left| \frac{\bar{X}-\mu_0}{\sigma_0/\sqrt{n}} \right| \right)\right] < \alpha \right) =$

$$= P_\mu\left(\Phi\left(\left| \frac{\bar{X}-\mu_0}{\sigma_0/\sqrt{n}} \right| \right) > 1 - \frac{\alpha}{2} \right) = P_\mu\left(\left| \frac{\bar{X}-\mu_0}{\sigma_0/\sqrt{n}} \right| > z_{1-\alpha/2} \right) =$$

$$= P_\mu\left(\frac{\bar{X}-\mu_0}{\sigma_0/\sqrt{n}} > z_{1-\frac{\alpha}{2}} \right) + P_\mu\left(\frac{\bar{X}-\mu_0}{\sigma_0/\sqrt{n}} < -z_{1-\frac{\alpha}{2}} \right) =$$

$$= P_\mu\left(\frac{\bar{X}-\mu}{\sigma_0/\sqrt{n}} > \frac{\mu_0-\mu}{\sigma_0/\sqrt{n}} + z_{1-\frac{\alpha}{2}}\right) + P_\mu\left(\frac{\bar{X}-\mu}{\sigma_0/\sqrt{n}} < \frac{\mu_0-\mu}{\sigma_0/\sqrt{n}} - z_{1-\frac{\alpha}{2}}\right) =$$

$$= 1 - \Phi\left(\frac{\mu_0-\mu}{\sigma_0/\sqrt{n}} + z_{1-\frac{\alpha}{2}}\right) + \Phi\left(\frac{\mu_0-\mu}{\sigma_0/\sqrt{n}} - z_{1-\frac{\alpha}{2}}\right)$$

Notice that $\beta(\mu) = \beta(\mu_0 + (\mu-\mu_0)) = \beta(\mu_0 - (\mu-\mu_0))$, so β is symmetric about μ_0.

Differentiating w.r.t. $\sqrt{n}$, find that

$$\left[\varphi\left(\frac{\mu_0-\mu}{\sigma_0/\sqrt{n}} - z_{1-\frac{\alpha}{2}}\right) - \varphi\left(\frac{\mu_0-\mu}{\sigma_0/\sqrt{n}} + z_{1-\frac{\alpha}{2}}\right)\right]\frac{\mu_0-\mu}{\sigma_0} \geq 0$$

this implies that $\beta(\mu)$ is $\uparrow$ as $n \uparrow$, so we need only solve $\beta(\mu_1) = \beta_0$ for n and determine a suitable sample size.

For example, when $\sigma_0 = 1$, $\alpha = 0.05$, $\beta_0 = 0.99$, $\mu_1 = \mu_0 \pm 0.1$, must find n s.t. $1 - \Phi(\sqrt{n}(0.1) + 1.96) + \Phi(\sqrt{n}(0.1) - 1.96) = 0.99$

$\leftarrow$ - -> may determine that $n = 785$ is the smallest possible sample size.

Also, the derivative w.r.t. μ is

$$\left[\varphi\left(\frac{\mu_0-\mu}{\sigma_0/\sqrt{n}} + z_{1-\frac{\alpha}{2}}\right) - \varphi\left(\frac{\mu_0-\mu}{\sigma_0/\sqrt{n}} - z_{1-\frac{\alpha}{2}}\right)\right]\frac{\sqrt{n}}{\sigma_0}$$

; it is >0 when $\mu > \mu_0$, <0 when $\mu < \mu_0$, and $=0$ when $\mu = \mu_0$.

Also, $\beta(\mu) \to 1$ as $\mu \to \pm\infty$. Thus, if we determine n such that the power is $\geq \beta_0$ for some μ_1, it will be $\geq \beta_0$ for any $|\mu_0 - \mu| \geq |\mu_0 - \mu_1|$

Ex. 6.3.19 Power function for the Bernoulli model.

Recall that $\beta(\theta) = P_\theta\left(2\left[1-\Phi\left(\left|\frac{\sqrt{n}(\bar{X}-\theta_0)}{\sqrt{\theta_0(1-\theta_0)}}\right|\right)\right] < \alpha\right)$

If n is large, & $\bar{X} \sim N(\theta, \theta(1-\theta)/n)$, we can compute the power as in Ex. 6.3.18 with $\sigma_0^2 = \theta(1-\theta)$.

Ex 6.3.20. Location-Scale Normal model.

For the 2-sided t-test,

$$\beta_n(\mu, \sigma^2) = P_{(\mu,\sigma^2)}\left(2\left[1 - G\left(\left|\frac{\bar{X}-\mu_0}{S/\sqrt{n}}\right|; n-1\right)\right] < \alpha\right) =$$

$$= P_{\mu,\sigma^2}\left(\left|\frac{\bar{X}-\mu_0}{S/\sqrt{n}}\right| > t_{1-\frac{\alpha}{2}}(n-1)\right)$$

We have to specify both μ and σ^2, then determine n s.t. $\beta_n(\mu, \sigma^2) \geq \beta_0$. One can e.g. use Monte-Carlo methods to approximate the cdf of $\left|\frac{\bar{X}-\mu_0}{S/\sqrt{n}}\right|$ for a variety of values of n, to determine an appropriate value