清华大学电子工程系

概率论与数理统计

2020 年春季学期

作业 10

王哲凡 2019011200 2020 年 4 月 29 日

10.1. *Solution.*

$$\overline{X} = 503.75, S \approx 6.4056$$

查表得：

$$t_{\frac{\alpha}{2}}(n-1) \approx 2.131$$

因此区间估计为：

$$\left(\overline{X} - t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}, \overline{X} + t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}\right) \approx (500.337, 507.163)$$

用平均重量作为估计，误差范围为 ± 3.413，是在 95% 置信意义下.

10.2. *Solution.*

$$\overline{X} = 1160, S \approx 111.5235$$

查表得：

$$t_{\alpha}(n-1) \approx 2.132$$

因此单侧置信下限为：

$$\overline{X} - t_{\alpha}(n-1)\frac{S}{\sqrt{n}} \approx 1053.667$$

10.3. (1). *Solution.*

$$\overline{X} = 91.73, \overline{Y} = 93.75, S \approx 1.99$$

查表得：

$$t_{\frac{\alpha}{2}}(n+m-2) \approx 2.0106$$

因此区间估计为：

$$\mu_1 - \mu_2 = \left(\overline{X} - \overline{Y}\right) \pm t_{\frac{\alpha}{2}}(n+m-2) S\sqrt{\frac{1}{n} + \frac{1}{m}} \approx -2.02 \pm 1.16$$

(2). *Solution.*

$$\frac{\left(\overline{X} - \overline{Y}\right) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_1^2}{n} + \dfrac{S_2^2}{m}}} \overset{\text{近似}}{\sim} N(0,1)$$

而：

$$z_{\frac{\alpha}{2}} \approx 1.96$$

因此区间估计为：

$$\mu_1-\mu_2=\left(\overline{X}-\overline{Y}\right)\pm z_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n}+\frac{S_2^2}{m}}\approx-2.02\pm1.12$$

(3). *Solution.* 由上述分析，可以看出两种催化剂的得率差值大约在 $-3\sim-1$ 之间，与其均值相比，可知两种催化局无显著差别.

10.4. 证明. 设 $\hat{\theta}=\max\{X_1,\cdots,X_n\}$，则：

$$\frac{\hat{\theta}}{\theta}=\max\left\{\frac{X_1}{\theta},\cdots,\frac{X_n}{\theta}\right\}$$

其中 $\frac{X_i}{\theta}\sim U(0,1)$，因此 $\dfrac{\hat{\theta}}{\theta}$ 是一个枢轴变量，且：

$$P\left(0<\frac{\hat{\theta}}{\theta}<1\right)=1$$

对于任意 $\alpha>0$，均存在 $x^*\in(0,1)$ 使得：

$$P\left(x^*<\frac{\hat{\theta}}{\theta}<1\right)\geq1-\alpha$$

转化即得：

$$P\left(\hat{\theta}<\theta<\frac{\hat{\theta}}{x^*}\right)\geq1-\alpha$$

因此取 $c_n=\dfrac{1}{x^*}$ 即可得到 θ 的 $(1-\alpha)$ 置信区间. □

10.5. *Solution.*

$$n\overline{X}\sim P(n\lambda)$$

设 $X\sim P(\lambda)$，则 X 的 cdf 为：

$$P(X\leq m|\lambda)=\sum_{i=0}^{m}\frac{\lambda^i}{i!}e^{-\lambda}$$

易得对于 λ：

$$f_m(\lambda)=P(X\leq m|\lambda)$$

是一个连续可导函数，且 $f_m(0^+)=1,f_m(+\infty)=0$.

下证明对于 $\forall\lambda\in\mathbb{R}$：

$$\sum_{t=m+1}^{\infty}\frac{\lambda^t}{t!}e^{-\lambda}=\frac{1}{m!}\int_0^{\lambda}t^me^{-\lambda}\,\mathrm{d}t$$

设：

$$g(\lambda)=\sum_{t=m+1}^{\infty}\frac{\lambda^t}{t!}e^{-\lambda}-\frac{1}{m!}\int_0^{\lambda}t^me^{-\lambda}\,\mathrm{d}t$$

由于第一项无穷级数易得一致收敛，则：

$$\begin{aligned} g'(\lambda) &= \sum_{t=m+1}^{\infty} t\frac{\lambda^{t-1}}{t!}e^{-\lambda} - \sum_{t=m+1}^{\infty} \frac{\lambda^t}{t!}e^{-\lambda} - \frac{1}{m!}\lambda^m e^{-\lambda} \\ &= \sum_{t=m}^{\infty} \frac{\lambda^t}{t!}e^{-\lambda} - \sum_{t=m+1}^{\infty} \frac{\lambda^t}{t!}e^{-\lambda} - \frac{1}{m!}e^{-\lambda} \\ &= 0 \end{aligned}$$

而 $g(0)=0$ 易得，因此 $g(\lambda)\equiv 0.$

考虑泊松总体的 cdf 为：

$$F(x;\lambda) = P(X \le x|\lambda) = f_m(\lambda), m = [x]$$

而在 x 或 m 确定时，λ 的 cdf 为（$\lambda > 0$）：

$$\frac{\dfrac{\lambda^m}{m!}e^{-\lambda}}{\displaystyle\int_0^{+\infty} \frac{t^m}{m!}e^{-t}\,\mathrm{d}t} = \frac{\lambda^m}{m!}e^{-\lambda}$$

故由上述结论，λ 的 pdf 即为：

$$\frac{1}{m!}\int_0^{\lambda} t^m e^{-t}\,\mathrm{d}t = 1 - F(x;\lambda)$$

而由于 $n\overline{X} \sim P(n\lambda)$，因此 λ 的置信度上下限 $\overline{\lambda}, \underline{\lambda}$ 可以由下式确定：

$$\frac{1}{m!}\int_0^{n\overline{\lambda}} t^m e^{-t}\,\mathrm{d}t = 1 - \frac{\alpha}{2}$$

$$\frac{1}{m!}\int_0^{n\underline{\lambda}} t^m e^{-t}\,\mathrm{d}t = \frac{\alpha}{2}$$

又因为：

$$\frac{1}{m!}\int_0^{\frac{x}{2}} t^m e^{-t}\,\mathrm{d}t = \frac{1}{2^{m+1}m!}\int_0^{x} t^m e^{-\frac{t}{2}}\,\mathrm{d}t$$

即为 $\chi^2(2m+2)$ 的 pdf，因此：

$$2n\underline{\lambda} = \chi^2_{\frac{\alpha}{2}}(2m+2), 2n\overline{\lambda} = \chi^2_{1-\frac{\alpha}{2}}(2m+2)$$

其中 $\chi^2_\alpha(2m+2)$ 为 $\chi^2(2m+2)$ 的下 $\alpha-$ 分位数.

因此 λ 的 $(1-\alpha)$ 置信区间为：

$$\left(\frac{\chi^2_{\frac{\alpha}{2}}(2m+2)}{2n}, \frac{\chi^2_{1-\frac{\alpha}{2}}(2m+2)}{2n}\right)$$

其中 $m = n\overline{X}$.

10.6. *Solution.* 第 i 个产品的合格情况 $X_i \sim B(p), i = 1, 2, \cdots, n = 100$，则：

$$\frac{\overline{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \overset{\text{近似}}{\sim} N(0, 1)$$

用 m_2 近似 $p(1-p) = Var(X_i)$，$m_2 = 0.24$，可得 95% 置信的区间估计为：

$$0.6 \pm 0.096$$

10.7. (1). *Solution.*

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

因此：

$$\begin{aligned} f(\overline{X}, \mu) &= f_M(\mu) f_X(\overline{X}|\mu) \\ &= \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}} \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} e^{-\frac{n(\overline{X}-\mu)^2}{2\sigma^2}} \\ &= \frac{\sqrt{n}}{2\pi\sigma\sigma_0} \exp\left(-\frac{1}{2}(A\mu^2 - B\mu + C)\right) \\ &\propto \frac{\sqrt{A}}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}A\left(\mu - \frac{B}{2A}\right)^2\right) \end{aligned}$$

故 μ 的后验分布为：

$$N\left(\frac{B}{2A}, \frac{1}{A}\right)$$

其中：

$$A = \frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}, B = \frac{2\mu_0}{\sigma_0^2} + \frac{2n\overline{X}}{\sigma^2}$$

即：

$$\mu \sim N\left(\frac{\sigma^2\mu_0 + n\sigma_0^2\overline{X}}{\sigma^2 + n\sigma_0^2}, \frac{\sigma^2\sigma_0^2}{\sigma^2 + n\sigma_0^2}\right)$$

因此 μ 的 $(1-\alpha)$ 置信区间为：

$$\left(\frac{\sigma^2\mu_0 + n\sigma_0^2\overline{X}}{\sigma^2 + n\sigma_0^2} - z_{\frac{\alpha}{2}}\frac{\sigma\sigma_0}{\sqrt{\sigma^2 + n\sigma_0^2}}, \frac{\sigma^2\mu_0 + n\sigma_0^2\overline{X}}{\sigma^2 + n\sigma_0^2} + z_{\frac{\alpha}{2}}\frac{\sigma\sigma_0}{\sqrt{\sigma^2 + n\sigma_0^2}}\right)$$

(2). *Solution.* 令 $\sigma_0 \to +\infty$，则置信区间变为：

$$\left(\overline{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$

与经典方法所求置信区间相同.

这是由于在当 $\sigma_0 \to +\infty$ 时，先验分布可以认为 $f_M(\mu) \propto 1$，也就是无先验信息可用时，只能靠样本，也就是经典方法.

10.8. 证明. 设：

$$g(x)=\begin{cases}p & x=1\\ 1-p & x=0\\ 0 & otherwise\end{cases}$$

设 $Y=X_1+\cdots+X_n$：

$$f(x_1,\cdots,x_n;p)=\prod_{i=1}^{n}g(x_i)=p^Y(1-p)^{n-Y}$$

$$\begin{aligned}f(x_1,\cdots,x_n|Y;p)&=\frac{f(x_1,\cdots,x_n;p)}{f(Y;p)}\\&=\frac{p^Y(1-p)^{n-Y}}{\binom{n}{Y}p^Y(1-p)^{n-Y}}=\frac{1}{\binom{n}{Y}}\end{aligned}$$

与 p 无关，因此 $Y=X_1+\cdots+X_n$ 的充分统计量. □

10.9. (1). 证明.

$$\begin{aligned}f(x_1,\cdots,x_n;\mu,\sigma^2)&=\prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(X_i-\mu)^2}{2\sigma^2}}\right)\\&=\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right)\exp\left(-\frac{\left(\overline{X}-\mu\right)^2}{\frac{2\sigma^2}{n}}\right)\\&=\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n\exp\left(-\frac{\left(\overline{X}-\mu\right)^2}{\frac{2\sigma^2}{n}}\right)\exp\left(-\frac{(n-1)S^2}{2\sigma^2}\right)\end{aligned}$$

由上次作业第八题的证明可知 $\overline{X}$ 与 S^2 独立.
而 $\overline{X}\sim N\left(\mu,\frac{\sigma^2}{n}\right),\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$，因此：

$$\begin{aligned}&f(x_1,\cdots,x_n|\overline{X},S^2;\mu,\sigma^2)\\=&\frac{f(x_1,\cdots,x_n;\mu,\sigma^2)}{f(\overline{X},S^2;\mu,\sigma^2)}\\=&\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{n-1}\frac{1}{\sqrt{n}}\exp\left(-\frac{(n-1)S^2}{2\sigma^2}\right)\frac{1}{f(S^2;\sigma^2)}\\=&\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{n-1}\frac{1}{\sqrt{n}}\frac{\sigma^2}{n-1}2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)\left(\frac{\sigma^2}{(n-1)S^2}\right)^{\frac{n-1}{2}-1}\\=&\left(\frac{1}{\sqrt{\pi}}\right)^{n-1}\frac{1}{(n-1)\sqrt{n}}\Gamma\left(\frac{n-1}{2}\right)\left(\frac{1}{(n-1)S^2}\right)^{\frac{n-1}{2}-1}\end{aligned}$$

与 (μ,σ^2) 无关，因此 $(\overline{X},S^2)$ 是 (μ,σ^2) 的充分统计量. □

(2). 证明. 设：

$$Y=\sum_{i=1}^{n}X_i=n\overline{X},Z=\sum_{i=1}^{n}X_i^2=n\overline{X}^2+(n-1)S^2$$

转化可得：

$$\overline{X}=\frac{Y}{n},S^2=\frac{Z}{n-1}-\frac{Y^2}{n(n-1)}$$

因此 (Y,Z) 的联合 pdf 为：

$$\begin{aligned}l(y,z;\mu,\sigma^2)&=|J|f\left(\frac{y}{n},\frac{z}{n-1}-\frac{y^2}{n(n-1)};\mu,\sigma^2\right)\\&=\frac{1}{\sqrt{2\pi n}\sigma}\exp\left(-\frac{n\left(\frac{y}{n}-\mu\right)^2}{2\sigma^2}\right)\frac{1}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)}\left(\frac{z-\frac{y}{n}}{\sigma^2}\right)^{\frac{n-1}{2}}\exp\left(-\frac{z-\frac{y}{n}}{2\sigma^2}\right)\end{aligned}$$

因此：

$$\begin{aligned}f(x_1,\cdots,x_n;\mu,\sigma^2)&=\prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(X_i-\mu)^2}{2\sigma^2}}\right)\\&=\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right)\exp\left(-\frac{\left(\overline{X}-\mu\right)^2}{\frac{2\sigma^2}{n}}\right)\\&=\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n\exp\left(-\frac{\left(\overline{X}-\mu\right)^2}{\frac{2\sigma^2}{n}}\right)\exp\left(-\frac{(n-1)S^2}{2\sigma^2}\right)\\&=\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n\exp\left(-\frac{\left(\frac{Y}{n}-\mu\right)^2}{\frac{2\sigma^2}{n}}\right)\exp\left(-\frac{Z-\frac{Y}{n}}{2\sigma^2}\right)\end{aligned}$$

故：

$$\begin{aligned}&f(x_1,\cdots,x_n|\overline{X},S^2;\mu,\sigma^2)\\=&\frac{f(x_1,\cdots,x_n;\mu,\sigma^2)}{l(Y,Z;\mu,\sigma^2)}\\=&\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n\exp\left(-\frac{\left(\frac{Y}{n}-\mu\right)^2}{\frac{2\sigma^2}{n}}\right)\exp\left(-\frac{Z-\frac{Y}{n}}{2\sigma^2}\right)\\&\left(\frac{1}{\sqrt{2\pi n}\sigma}\exp\left(-\frac{n\left(\frac{Y}{n}-\mu\right)^2}{2\sigma^2}\right)\frac{1}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)}\left(\frac{Z-\frac{Y}{n}}{\sigma^2}\right)^{\frac{n-1}{2}}\exp\left(-\frac{Z-\frac{Y}{n}}{2\sigma^2}\right)\right)^{-1}\\=&\sqrt{2\pi n}\sigma\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n\frac{1}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)}\left(\frac{Z-\frac{Y}{n}}{\sigma^2}\right)^{\frac{n-1}{2}}\\=&\sqrt{n}\left(\frac{1}{\sqrt{\pi}}\right)^{n-1}\frac{1}{\Gamma\left(\frac{n-1}{2}\right)}\left(Z-\frac{Y}{n}\right)^{\frac{n-1}{2}}\end{aligned}$$

与 (μ,σ^2) 无关，因此 $(Y,Z)=\left(\sum\limits_{i=1}^{n}X_i,\sum\limits_{i=1}^{n}X_i^2\right)$ 是 (μ,σ^2) 的充分统计量. □

10.10. (1). *Solution.*

$$\theta\overset{\text{近似}}{\sim}N(\overline{\theta},V_{boot})\approx N(149,224)$$

因此 θ 的 95% 置信区间估计为：

$$(120, 179)$$

(2). *Solution.* 由于：

$$\frac{\mu - \overline{X}}{\frac{1}{n}} \sim N(0,1) \Rightarrow \mu \sim N\left(\overline{X}, \frac{1}{100}\right)$$

根据 μ 的 95% 置信区间估计：

$$\mu \approx \overline{X} \pm 0.196$$

则对应取 exp 得 θ 的 95% 置信区间估计为：

$$\left(e^{\overline{X}-0.196}, e^{\overline{X}+0.196}\right)$$

如果假设 $\overline{X} \approx 5$，则对应区间估计约为：

$$(122, 181)$$

由于 θ, μ 对应区间的概率相等，故其 95% 置信区间也就对应.