统计预习检测

TechX

June 2022

1 补充资料

概率论与数理统计网课; bilibili

教材: *Probability and Statistical Inference*: Google 云盘

2 检测

2.1 条件概率

一家保险公司出售几种类型的保单，包括汽车保单和房主保单。设 A 1 是那些只有汽车保单的人，A2 是那些只有房主保单的人，A 3 是那些既有汽车保单又有房主保单的人（但没有其他保单）。对于一个从公司保单持有人中随机抽取的人，假设 $P(A_1) = 0.3$, $P(A_2) = 0.2$, $P(A_3) = 0.2$。此外，让 B 是这个人至少会续签其中一份保单的事件。根据过去的经验，我们假设条件概率 $P(B|A_1) = 0.6$, $P(B|A_2) = 0.7$, $P(B|A_3) = 0.8$。考虑到随机选择的人有一份汽车或房主保单，那么这个人至少会更新其中一份保单的条件概率是多少?

2.2 独立事件

骰子 A 有两个面是橙色，四个面是蓝色，骰子 B 有三个面是橙色，三个面是蓝色，骰子 C 有四个面是橙色，两个面是蓝色。所有这些都是公平的骰子。如果掷出这三个骰子，求三个骰子中正好有两个是橙色的概率。

2.3 贝叶斯定理

在医院的急诊室，病人被分类，其中 20% 是危重病人，30% 是严重的，50% 是稳定的。在危重的病人中，30% 死亡；在严重的病人中，10% 死亡；而在稳定的病人中，1% 死亡。考虑到有病人死亡，那么该病人被列为危重的条件概率是多少?

2.4 随机变量

对于以下每一项，确定常数 c，使 $f(x)$ 满足作为随机变量 X 的 pmf 的条件。

- (a) $f(x) = x/c, \quad x = 1, 2, 3, 4.$
- (b) $f(x) = cx, \quad x = 1, 2, 3, ..., 10.$
- (c) $f(x) = c(1/4)^x, \quad x = 1, 2, 3, ...$
- (d) $f(x) = c(x + 1)^2, \quad x = 0, 1, 2, 3.$
- (e) $f(x) = x/c, \quad x = 1, 2, 3, ..., n.$
- (f) $f(x) = \frac{c}{(x+1)(x+2)}, \quad x = 0, 1, 2, 3, ...$
  Hint: $f(x) = \frac{1}{x(x+1)} = \frac{1}{x} - \frac{1}{(x+1)}$

2.5 均匀分布

对于连续型随机变量 $X \sim U(a, b)$，证明 $E(x) = \frac{a+b}{2}$，$Var(x) = \frac{(a-b)^2}{12}$。

2.6 正态分布

[矩母函数 $M(t) = E(e^{tx})$] 尝试推导正态分布 $X \sim N(\mu, \sigma^2)$ 的矩母函数。
Hint:

$$\int \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{(x-\mu)^2}{2\sigma^2})dx = 1$$

2.7 假设检验

在美国，婴儿的平均出生体重是 3315 克。假设 X 是随机选择的一个婴儿的出生体重(克)。X 的分布是 $N(\mu, \sigma^2)$，其中 μ 和 σ^2 是未知的。我们将用 n = 30 个随机选择的婴儿来检验无效假设 $H_0 : \mu = 3315$ 与备选假设 $H_1 : \mu < 3315$。

(a) 当 n=30, 随机样本具有统计量: $\overline{x} = 3189$ 和 $S = 488$, 你的结论是什么 (显著性水平 $\alpha = 0.05$)?
(b) 你的测试的 $p - value$ 大约是多少?

2.8 线性回归

基于最小二乘法 (least square), 推导线性回归的参数 α 和 β

3 答案

3.1 条件概率

希望的概率是:

$$\begin{aligned}
P(B|A_1 \cup A_2 \cup A_3) &= \frac{P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B)}{P(A_1) + P(A_2) + P(A_3)} \\
&= \frac{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)}{P(A_1) + P(A_2) + P(A_3)} \\
&= \frac{0.3 \cdot 0.6 + 0.2 \cdot 0.6 + 0.2 \cdot 0.7}{0.3 + 0.2 + 0.2} \\
&= \frac{0.48}{0.70} = 0.686
\end{aligned}$$

3.2 独立事件

掷骰子 A，B，和 C 为独立事件，即:

$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$$

P(A= 橙色)=$\frac{2}{6} = \frac{1}{3}$ P(B= 橙色)=$\frac{3}{6} = \frac{1}{2}$ P(C= 橙色)=$\frac{4}{6} = \frac{2}{3}$

D: 掷三个骰子其中两个为橙色

$P(D) = P(A \cap B \cap C') + P(A \cap B' \cap C) + P(A' \cap B \cap C) = \frac{1}{18} + \frac{1}{9} + \frac{2}{9} = \frac{7}{18}$

3.3 贝叶斯定理

A: 危重病人 B: 严重病人 C: 稳定病人 D: 病人死亡

病人死亡并被列为为重的条件概率: P(B|D)

$P(A) = 20\% \quad P(B) = 30\% \quad P(C) = 50\%$

$P(A \cap D) = 30\% \quad P(B \cap D) = 10\% \quad P(C \cap D) = 1\%$

$P(B|D) = \dfrac{P(B \cap D)}{P(D)} = \dfrac{P(B \cap D)}{P(D \cap (A \cup B \cup C))} = \dfrac{P(B \cap D)}{\sum_{I \in \{A,B,C\}} P(I \cap D)}$

$P(D \cap (A \cup B \cup C)) = P((D \cap A) \cup (D \cap B) \cup (D \cap C)) = P(D \cap A) + P(D \cap B) + P(D \cap C)$

$P(A \cap B \cap C) = 0$

$P(B|D) = \dfrac{0.1}{0.3 + 0.1 + 0.01} = 24.4\%$

3.4 随机变量

借助定义 $\sum_{x \in S} f(x) = 1$

(a) c = 10

(b) c = 1/55

(c) c = 3

(d) c = 1/30

(e) c = 2n(n + 1)

(f) c = 1 ($f(x) = c(\frac{1}{(x+1)} - \frac{1}{(x+2)})$)

3.5 均匀分布

$$f(x) = \frac{1}{(b-a)}$$

$$
\begin{aligned}
E(X) &= \int_a^b xf(x)dx = \int_a^b \frac{x}{(b-a)}dx \\
&= [\frac{x^2}{2(b-a)}]_a^b = \frac{b^2}{2(b-a)} - \frac{a^2}{2(b-a)} \\
&= \frac{a+b}{2} \\
E(X^2) &= \int_a^b x^2 f(x)dx = \int_a^b \frac{x^2}{(b-a)}dx \\
&= \frac{b^3}{3(b-a)} - \frac{a^3}{3(b-a)} \\
&= \frac{a^2+ab+b^2}{3} \\
Var(X) &= E(X^2) - (E[X])^2 \\
&= \frac{a^2+ab+b^2}{3} - \frac{a^2+2ab+b^2}{4} \\
&= \frac{a^2-2ab+b^2}{12} = \frac{(a-b)^2}{12}
\end{aligned}
$$

3.6 正态分布

$$
\begin{aligned}
E(e^{tx}) &= \int \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{(x-\mu)^2}{2\sigma^2}) \cdot exp(tx)dx \\
&= \int \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{(x-\mu)^2}{2\sigma^2} + tx)dx \\
&= \int \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{1}{2\sigma^2}[x^2 - 2(\mu+t\sigma^2)x + \mu^2 + (\mu+t\sigma^2)^2 - (\mu+t\sigma^2)^2])dx \\
&= \int \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{1}{2\sigma^2}(x-\mu-t\sigma^2)^2) exp(\frac{1}{2\sigma^2}(t^2\sigma^4 + 2\mu t\sigma^2))dx \\
&= exp(\frac{t^2\sigma^2+2\mu t}{2})
\end{aligned}
$$

$since\ \int \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{1}{2\sigma^2}(x-\mu-t\sigma)^2)dx = 1,\ X \sim N(\mu+t\sigma^2, \sigma^2)$

3.7 假设实验

(a) $\alpha = 0.05,\ H_0 : \mu = 3315$

构造 $T=\frac{\bar{x}-\mu}{S/\sqrt{n}}\sim t(29)$

$t=\frac{3189-3315}{488/\sqrt{30}}=-1.414$

$-t_{\alpha}(n-1)=-t_{0.05}(29)=-1.699<t$

因此，在显著性程度为 0.05 时，我们接受无效假设,$\mu=3315$

(b) $p-value: P(T<t|\mu=3315)=F_{t(29)}(t)=0.084$(可用 excel 函数计算或近似估计)

3.8 线性回归

$$\hat{y}=\beta x+\alpha$$

最小二乘法: 使残差和最小, 关于参数 α 和 β 的偏导数为 0

$$min\sum(y_i-\beta x_i-\alpha)^2$$

$$\frac{\partial}{\partial\alpha}=\sum(y_i-\alpha)=0 \tag{1}$$

$$\frac{\partial}{\partial\beta}=\sum(2\beta x_i^2+2\alpha x_i-2x_iy_i)=0 \tag{2}$$

联立得:

$$\begin{aligned}
&\hat{\alpha}=\frac{\sum y_i}{n}=\bar{y}\\
&\hat{\beta}=\frac{\sum y_ix_i-\sum\bar{y}x_i}{\sum x_i^2}=\frac{\sum(y_i-\bar{y})x_i}{\sum x_i^2}\\
&\hat{y}=\hat{\beta}x+\hat{\alpha}
\end{aligned}$$