Midterm Review

吴梦轩

1 随机变量

1.1 重要分布

1.1.1 二项分布

二项分布的符号表示为 $X \sim b(n,p)$，其频率函数为：

$$P\{X=x\}=C_n^x p^x(1-p)^{n-x}=\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$$

意为：n 次独立重复试验中，事件 A 发生 x 次的概率。

多项分布：

多项分布的符号表示为 $X \sim m(n,p_1,p_2,\cdots,p_k)$，其频率函数为：

$$P\{X_1=x_1,X_2=x_2,\cdots,X_k=x_k\}=\frac{n!}{x_1!x_2!\cdots x_k!}p_1^{x_1}p_2^{x_2}\cdots p_k^{x_k}\quad\left(\sum_{i=1}^k p_i=1,\sum_{i=1}^k x_i=n\right)$$

意为：n 次独立重复试验中，事件 $A_1,A_2,\cdots,A_k$ 分别发生 $x_1,x_2,\cdots,x_k$ 次的概率。

1.1.2 泊松分布

泊松分布的符号表示为 $X \sim P(\lambda)$ 或 $X \sim \pi(\lambda)$，其频率函数为：

$$P\{X=x\}=\frac{\lambda^x}{x!}e^{-\lambda}$$

意为：单位时间（或单位面积）内随机事件发生 x 次的概率。

泊松定理：

设 $X \sim b(n,p)$，当 $n\to\infty,p\to 0$，使得 $np=\lambda$ 不变，则有：

$$\lim_{n\to\infty}P\{X=x\}=\lim_{n\to\infty}C_n^x p^x(1-p)^{n-x}=\frac{\lambda^x}{x!}e^{-\lambda}$$

泊松流：

记在时间 $(0,t]$ 内随机事件发生的次数为 $N(t)$，则 $N(t) \sim P(\lambda t)$，称 $N(t)$ 为泊松流。

$$P\{N(t)=n\} = \frac{(\lambda t)^n}{n!}e^{-\lambda t}$$

1.1.3 均匀分布

均匀分布的符号表示为 $X \sim U(a,b)$，其频率函数为:

$$P\{X=x\} = \begin{cases} \frac{1}{b-a}, & a<x<b \\ 0, & \text{otherwise} \end{cases}$$

其分布函数为:

$$F(x) = \begin{cases} 0, & x \leqslant a \\ \frac{x-a}{b-a}, & a<x<b \\ 1, & x \geqslant b \end{cases}$$

1.1.4 正态分布

正态分布的符号表示为 $X \sim N(\mu,\sigma^2)$，其频率函数为:

$$P\{X=x\} = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

特别的，当 $\mu=0,\sigma=1$ 时，称为标准正态分布，记为 $X \sim N(0,1)$。标准正态分布的分布函数记为 $\Phi(x)$，其频率函数记为 $\varphi(x)$。

分布函数的积分过程:

$$\begin{aligned} F(+\infty) &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\mathrm{d}x \end{aligned}$$

此时令 $t=\frac{x-\mu}{\sigma}$，则有:

$$\begin{aligned} F(+\infty) &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}}\mathrm{d}t \\ &= \frac{1}{\sqrt{2\pi}}\sqrt{2\pi} \\ &= 1 \end{aligned}$$

规范化： 证明若 $X \sim N(\mu, \sigma^2)$，则 $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$。

$$\begin{aligned}
F_Z(z) &= P\{Z \leqslant z\} \\
&= P\left\{\frac{X-\mu}{\sigma} \leqslant z\right\} \\
&= P\{X \leqslant \sigma z + \mu\} \\
&= F_X(\sigma z + \mu) \\
&= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\sigma z+\mu} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x
\end{aligned}$$

此时令 $t = \frac{x-\mu}{\sigma}$，则有：

$$\begin{aligned}
F_Z(z) &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\sigma z+\mu} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-\frac{t^2}{2}} \mathrm{d}t \\
&= \Phi(z)
\end{aligned}$$

1.1.5 指数分布

指数分布的符号表示为 $X \sim EXP(\lambda)$，其频率函数为：

$$P\{X = x\} = \begin{cases} \lambda e^{-\lambda x}, & x \geqslant 0 \\ 0, & x < 0 \end{cases}$$

其分布函数为：

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geqslant 0 \\ 0, & x < 0 \end{cases}$$

指数分布与泊松流：

设 $N(t)$ 为泊松流，$N(t) \sim P(\lambda t)$。记 X 为第一个事件发生的时间，则 $P\{X > t\} = P\{N(t) = 0\} = e^{-\lambda t}$。因此 $X \sim EXP(\lambda)$。

指数分布的无记忆性：

设 $X \sim EXP(\lambda)$，则对任意 $s, t > 0$，有：

$$\begin{aligned}
P\{X > s+t | X > s\} &= \frac{P\{X > s+t, X > s\}}{P\{X > s\}} \\
&= \frac{P\{X > s+t\}}{P\{X > s\}} \\
&= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} \\
&= e^{-\lambda t} \\
&= P\{X > t\}
\end{aligned}$$

1.2 频率（密度）函数与分布函数

频率函数的基本性质（本质特征）:

- $P\{X = x\} \geqslant 0$
- $\sum_{x \in X} P\{X = x\} = 1$

满足以上两个性质的数列必定是一个离散型随机变量的频率函数。
密度函数的基本性质（本质特征）:

- $f(x) \geqslant 0$
- $\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1$

满足以上两个性质的函数必定是一个连续型随机变量的密度函数。
分布函数的基本性质（本质特征）:

- $F(x)$ 是一个单调不减函数
- $\lim_{x\to-\infty} F(x) = 0, \lim_{x\to+\infty} F(x) = 1$
- $F(x)$ 是右连续的，即 $\lim_{x\to x_0^+} F(x) = F(x_0)$

满足以上三个性质的函数必定是一个随机变量的分布函数。

1.3 分位数

分位数：设 X 是一个连续型随机变量，$F(x)$ 是其分布函数，对于 $0 < p < 1$，若实数 x_p 使得 $F(x_p) = p$，则称 x_p 为 X 的 p 分位数。特殊的，当 $p = 0.5$ 时，称为中位数。当 $p = 0.25$ 时，称为下四分之一位数。当 $p = 0.75$ 时，称为上四分之一位数。

1.4 随机变量的函数的分布

令 $Y = g(X)$，当 Y 是严格单调函数时，有：

$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y}\right|$$

正态分布的线性组合以及线性函数的分布仍然是正态分布。假设有 n 个随机变量 $X_1, X_2, \cdots, X_n$，且 $X_i \sim N(\mu_i, \sigma_i^2)$，则有：

$$\sum_{i=1}^{n} a_i X_i + b \sim N\left(\sum_{i=1}^{n} a_i\mu_i + b, \sum_{i=1}^{n} a_i^2\sigma_i^2\right)$$

1.5 二维随机变量

1.5.1 联合分布函数与边缘分布函数

根据 Farlie-Morgenstern 定理，对于给定的两个一维随机变量 X,Y，可以证明只要 $|a|\leqslant 1$，就有：

$$H(x,y)=F(x)G(y)1+a[1-F(x)][1-G(y)]$$

是二元连续型分布函数。

换言之，给定边际分布，可以构造出无限多个联合分布。

1.5.2 连接函数

将使得边缘分布为均匀分布的联合分布函数称为连接函数。定义联合分布为：

$$F_{XY}(x,y)=C(F_X(x),F_Y(y))$$

定义密度函数为：

$$f_{XY}(x,y)=c(F_X(x),F_Y(y))f_X(x)f_Y(y)$$

1.5.3 二维正态分布

二维正态分布记为 $X\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，其联合密度函数为：

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}$$

对于其边际分布，有：

$$X\sim N(\mu_1,\sigma_1^2),\ Y\sim N(\mu_2,\sigma_2^2)$$

证明如下：

$$\begin{aligned}
f_X(x)&=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y\\
&=\int_{-\infty}^{+\infty}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}\mathrm{d}y\\
&=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right\}\cdot\\
&\quad\int_{-\infty}^{+\infty}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(y-\mu_2)^2}{\sigma_2^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+p^2\frac{(x-\mu_1)^2}{\sigma_1^2}\right]\right\}\mathrm{d}y\\
&=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right\}\cdot\\
&\quad\int_{-\infty}^{+\infty}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{y-\mu_2}{\sigma_2}-\rho\frac{x-\mu_1}{\sigma_1}\right]^2\right\}\mathrm{d}y
\end{aligned}$$

令 $t=\frac{1}{\sqrt{1-\rho^2}}\left(\frac{y-\mu_2}{\sigma_2}-\rho\frac{x-\mu_1}{\sigma_1}\right)$，则有：

$$\begin{aligned}f_X(x)=&\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right\}\cdot\\&\int_{-\infty}^{+\infty}\exp\left\{-\frac{t^2}{2}\right\}\frac{1}{\sigma_2\sqrt{1-\rho^2}}\mathrm{d}t\\=&\frac{1}{\sqrt{2\pi}\sigma_1}\exp\left\{-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right\}\\=&N(\mu_1,\sigma_1^2)\end{aligned}$$

注意：边际分布为正态分布时，联合分布不一定为二维正态分布。

1.5.4 二维独立性

设 X,Y 为二维随机变量，若对于任意 x,y，有：

$$F(x,y)=F_X(x)F_Y(y)$$

则称 X,Y 相互独立。

对于离散型随机变量，有：

$$P\{X=x,Y=y\}=P\{X=x\}P\{Y=y\}$$

对于连续型随机变量，有：

$$f(x,y)=f_X(x)f_Y(y)$$

特别的，二维正态分布的独立性等价于 $\rho=0$。

两组独立的数据 $(X_1,X_2,\cdots,X_n)$ 和 $(Y_1,Y_2,\cdots,Y_m)$，其函数 $h(X_1,X_2,\cdots,X_n)$ 和 $g(Y_1,Y_2,\cdots,Y_m)$ 相互独立。

1.5.5 条件分布

设 X,Y 为二维随机变量，$f(x,y)$ 为其联合密度函数，$f_X(x)$ 为 X 的边际密度函数，若 $f_X(x)>0$，则称：

$$f_{Y|X}(y|x)=\frac{f(x,y)}{f_X(x)}$$

为 Y 在 $X=x$ 的条件密度函数。

二维随机变量的全概率公式：

$$f_Y(y)=\int_{-\infty}^{+\infty}f_{Y|X}(y|x)f_X(x)\mathrm{d}x$$

特别的，对于二维正态分布，指定 X 的条件下 Y 的条件分布仍然是正态分布。准确的说，若 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，则有：

$$f_{Y|X}(y|x)\sim N\left(\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1),\sigma_2^2(1-\rho^2)\right)$$

1.5.6 联合分布随机变量的函数

设 X, Y 为二维随机变量，$Z = f(X, Y)$，则有：

$$F_Z(z) = P\{Z \leqslant z\} = P\{f(X, Y) \leqslant z\} = \iint\limits_{f(x,y)\leqslant z} f(x, y)\mathrm{d}x\mathrm{d}y$$

$Z = X + Y$ **的分布：**

$$\begin{aligned} P\{Z \leqslant z\} &= P\{X + Y \leqslant z\} \\ &= P\{X \leqslant z - Y\} \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{z-y} f(x, y)\mathrm{d}x\mathrm{d}y \end{aligned}$$

令 $x = u - y$，则有：

$$\begin{aligned} P\{Z \leqslant z\} &= \int_{-\infty}^{+\infty}\int_{-\infty}^{z} f(u - y, y)\mathrm{d}u\mathrm{d}y \\ f_Z(z) &= \int_{-\infty}^{\infty} f(z - y, y)\mathrm{d}y \\ &= \int_{-\infty}^{\infty} f(x, z - x)\mathrm{d}x \end{aligned}$$

特别的，当 X, Y 相互独立时，有：

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z - x)\mathrm{d}x = \int_{-\infty}^{\infty} f_X(z - y)f_Y(y)\mathrm{d}y$$

特别的，当 $X \sim P(\lambda_1), Y \sim P(\lambda_2)$ 时，$Z = X + Y \sim P(\lambda_1 + \lambda_2)$。

$Z = \frac{X}{Y}$ **的分布：**

同理，通过换元法，有：

$$f_Z(z) = \int_{-\infty}^{\infty} |y|f_X(zy, y)\mathrm{d}y = \int_{-\infty}^{\infty} |x|f_X(x, xz)\mathrm{d}x$$

两个随机变量变换的分布：

如果两个随机变量的联合分布为二维正态分布，则他们的非奇异线性变换的分布仍然是二维正态分布。

1.6 极值

$$
\begin{aligned}
F_{\max}(z) &= P\{X \leqslant z, Y \leqslant z\} \\
&= P\{X \leqslant z\}P\{Y \leqslant z\} \\
&= F_X(z)F_Y(z) \\
F_{\min}(z) &= 1 - P\{X > z, Y > z\} \\
&= 1 - P\{X > z\}P\{Y > z\} \\
&= 1 - [1 - F_X(z)][1 - F_Y(z)]
\end{aligned}
$$

特别的，如果 $X \sim EXP(\lambda), Y \sim EXP(\mu)$，则 $Z = \min\{X, Y\} \sim EXP(\lambda + \mu)$。

1.7 顺序统计量

$X_{(k)}$ 的密度函数为：

$$
f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!}[F(x)]^{k-1}[1-F(x)]^{n-k}f(x)
$$

$(X_{(1)}, X_{(n)})$ 的联合密度函数为：

$$
f_{X_{(1)},X_{(n)}}(x, y) = n(n-1)[F(y) - F(x)]^{n-2}f(x)f(y)
$$

2 古典概型

2.1 集合

试验的样本点记为 ω，样本空间记为 Ω，样本空间的子集称为事件。

当事件 A 与事件 B 不可能同时发生时，称事件 A 与事件 B 互不相容，或互斥，或互不相交，记为 $A \cap B = \varnothing$。

加法公式：

$$
P(A \cup B) = P(A) + P(B) - P(AB)
$$

2.1.1 计数方法

选排列：从 n 个不同元素中任取 m 个元素，按照一定的顺序排成一列，称为从 n 个不同元素中选取 m 个元素的排列，记为 $A_n^m = \frac{n!}{(n-m)!}$。特别的，当 $m = n$ 时，称为全排列，记为 $A_n^n = n!$。

组合：从 n 个不同元素中任取 m 个元素，不考虑顺序，称为从 n 个不同元素中选取 m 个元素的组合，记为 $C_n^m = \frac{A_n^m}{m!} = \frac{n!}{m!(n-m)!}$。特别的，将 n 个不同元素分成 r 组，

使得第 i 组有 n_i 个元素，且 $\sum_{i=1}^{r} n_i = n$，则称为将 n 个不同元素分成 r 组的组合，记为 $C_n^{n_1,n_2,\cdots,n_r} = \frac{n!}{n_1!n_2!\cdots n_r!}$。

将 n 个相同的球放入 m 个不同的盒子中，每个盒子中至少有一个球的方法数为 C_{n-1}^{m-1}。（相当于在 $n-1$ 个间隔中选取 $m-1$ 个间隔放入隔板）将 n 个相同的球放入 m 个不同的盒子中，盒子可以为空的方法数为 C_{n+m-1}^{m-1}。（相当于将 $n+m$ 个相同的球放入 m 个不同的盒子中，每个盒子中至少有一个球）

2.2 全概率公式与贝叶斯公式

全概率公式：设 $B_1, B_2, \cdots, B_n$ 为样本空间 Ω 的一个划分，即 $B_i \cap B_j = \varnothing(i \neq j), \bigcup_{i=1}^{n} B_i = \Omega$，则对任一事件 A，有：

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

贝叶斯公式：设 $B_1, B_2, \cdots, B_n$ 为样本空间 Ω 的一个划分，即 $B_i \cap B_j = \varnothing(i \neq j), \bigcup_{i=1}^{n} B_i = \Omega$，则对任一事件 A，有：

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)}$$

2.3 独立性

设 A, B 为两事件，若 $P(AB) = P(A)P(B)$，则称事件 A 与事件 B 相互独立。

独立不同于互不相容，独立是指两事件发生的概率互不影响，互不相容是指两事件不能同时发生。

两两独立不能推出多个事件相互独立。独立也没有传递性，即 A, B 独立，B, C 独立，不能推出 A, C 独立。