1.(P204,8) **假设总体X服从$0-1$分布$B(1,p)$，其中p为未知参数，$(X_1,X_2,\ldots,X_5)$为从此总体中抽取的简单样本.**

(1)写出样本空间和抽样分布

(2)指出$X_1+X_2,\min_{1\leqslant i\leqslant 5}X_i,X_5+2p,X_5-E(X_1),\frac{(X_5-X_1)^2}{Var(X_1)}$哪些是统计量，哪些不是，为什么?

解:

(1)样本空间：$\{(x_1,x_2,x_3,x_4,x_5)|x_i=0或1,i=1,2,3,4,5\}$

因为每次抽取互相独立，抽样分布：

$$\begin{aligned}&P(X_1=x_1,X_2=x_2,X_3=x_3,X_4=x_4,X_5=x_5)\\=&P(X_1=x_1)P(X_2=x_2)P(X_3=x_3)P(X_4=x_4)P(X_5=x_5)\\=&(\prod_{i=1}^{k}P(X=1))(\prod_{i=1}^{5-k}P(X=0))\\=&p^k(1-p)^{5-k},\qquad\sum_{i=1}^{5}x_i=k,0\leqslant k\leqslant 5\end{aligned}$$

(2)$X_1+X_2,\min_{1\leqslant i\leqslant 5}X_i$是统计量,而$X_5+2p,X_5-E(X_1),\frac{(X_5-X_1)^2}{Var(X_1)}$不是，因为后者中有不由$X_1,X_2,X_3,X_4,X_5$决定的$p,E(X_1),Var(X_1)$

2.(P205,15) **设X_1,X_2,X_3,X_4是来自正态总体$N(0,2^2)$的简单随机样本，令$T=a(X_1-2X_2)^2+b(3X_3-4X_4)^2$.试求$a,b$使统计量$T$服从$\chi^2$分布.**

解:

由题，因为$X_1,X_2,X_3,X_4\sim N(0,2^2)$，所以有$X_1-2X_2\sim N(0,20),3X_3-4X_4\sim N(0,100)$，所以有$\frac{X_1-2X_2}{\sqrt{20}}\sim N(0,1),\frac{3X_3-4X_4}{10}\sim N(0,1)$，则要使$T$服从$\chi^2$分布，则使$a=\frac{1}{20},b=\frac{1}{100}$即可.

3.(P205,18) **设$X_1,X_2,\ldots,X_n$为从下列总体中抽取的简单样本:**

(1)正态总体$N(\mu,\sigma^2)$;

(2)参数为λ的泊松总体;

(3)参数为λ的指数分布;

试求样本均值$\bar{X}$的分布.

解:

(1)由定理，有$\bar{X}\sim N(\mu,\frac{\sigma^2}{n})$

(2)$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$，则由泊松分布的可加性有$X_1+X_2+\ldots+X_n\sim P(n\lambda),P(\sum X_i=k)=\frac{(n\lambda)^k}{k!}e^{-n\lambda}$，所以$P(\bar{X}=k)=P(\sum X_i=nk)=\frac{(n\lambda)^{nk}}{(nk)!}e^{-n\lambda}$

(3)$f(x)=\lambda e^{-\lambda x}I_{[0,+\infty)}$，则$X_1+X_2$服从的分布为

$$P(X_1+X_2\leqslant z)=\int_0^z\int_0^{z-x}\lambda e^{-\lambda x}\lambda e^{-\lambda y}\mathrm{d}x\mathrm{d}y=1-(1+\lambda z)e^{-\lambda z}$$

有$f_{X_1+X_2}(x)=\lambda^2xe^{-\lambda x}=\frac{\lambda^2}{\Gamma(2)}xe^{-\lambda x}$ 若$f_{X_1+X_2+\ldots+X_n}(x)=\frac{\lambda^n}{\Gamma(n)}x^{n-1}e^{-\lambda x}$有

$$P(X_1+X_2+\ldots+X_n+X_{n+1}\leqslant z)=\int_0^z\int_0^{z-x}\frac{\lambda^n}{\Gamma(n)}x^{n-1}e^{-\lambda x}\lambda e^{-\lambda y}\mathrm{d}x\mathrm{d}y=1-(1+\lambda z+\ldots+\frac{\lambda^nz^n}{n!})e^{-\lambda z}$$

则$f_{X_1+X_2+\ldots+X_n}(x)=\frac{\lambda^{n+1}}{\Gamma(n+1)}x^ne^{-\lambda x}$

$f_{\bar{X}}(x)=\frac{\lambda^{n+1}}{\Gamma(n+1)}(nx)^ne^{-\lambda nx}$

4.(P205,19) **设$(X_1,X_2,\ldots,X_n)$是从$0-1$分布$B(1,p)$中抽取的简单样本，$0<p<1$，记$\bar{X}$为样本均值，求$S_n^2=\sum_{i=1}^{n}(X_i-\bar{X})^2/n$的期望.**

解:

$$
\begin{aligned}
ES_n^2 &= E(\frac{\sum_{i=1}^n (X_i-\bar{X})^2}{n}) \\
&= E\frac{\sum_{i=1}^n X_i^2}{n} - 2E\frac{\bar{X}\sum_{i=1}^n X_i}{n} + E\frac{\sum_{i=1}^n \bar{X}^2}{n} \\
&= \frac{E\sum_{i=1}^n X_i^2}{n} - 2E\bar{X}^2 + E\bar{X}^2 \\
&= \frac{E\sum_{i=1}^n X_i^2}{n} - E\bar{X}^2 \\
&= \frac{E\sum_{i=1}^n X_i^2}{n} - \frac{E(\sum_{i=1}^n X_i)^2}{n^2}
\end{aligned}
$$

由题，有$P(\sum_{i=1}^n X_i^2 = k) = C_n^k p^k (1-p)^{n-k}, E\sum_{i=1}^n X_i^2 = \sum_{k=0}^n kC_n^k p^k (1-p)^{n-k} = np\sum_{k=1}^n C_{n-1}^{k-1} p^{k-1}(1-p)^{n-k} = np$
$P(\sum_{i=1}^n X_i = k) = C_n^k p^k (1-p)^{n-k}$
$E(\sum_{i=1}^n X_i)^2 = \sum_{k=0}^n k^2 C_n^k p^k (1-p)^{n-k} = np\sum_{k=1}^n kC_{n-1}^{k-1}p^{k-1}(1-p)^{n-k} = np((n-1)p\sum_{k=2}^n C_{n-2}^{k-2}p^{k-2}(1-p)^{n-k}+1) = n(n-1)p^2+np$
则$ES_n^2 = np/n - [n(n-1)p^2+np]/n^2 = \frac{n-1}{n}p(1-p)$

> 5.(P206,20) **设$(X_1, X_2, \dots, X_n)$为来自正态总体$N(\mu, \sigma^2)$的一个简单随机样本，$\bar{X}$和S_n^2分别表示样本均值和样本方差，又设$X_{n+1} \sim N(\mu, \sigma^2)$且与$X_1, X_2, \dots, X_n$独立，试求统计量$\sqrt{\frac{n}{n+1}}(X_{n+1}-\bar{X})/S_n$的分布.**

解：易知有$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}), X_{n+1}-\bar{X} \sim N(0, \frac{n+1}{n}\sigma^2), \sqrt{\frac{n}{n+1}}(X_{n+1}-\bar{X}) \sim N(0, \sigma^2)$

而由定理有$\frac{n-1}{\sigma^2}S_n^2 \sim \chi_{n-1}^2$，且$\frac{n-1}{\sigma^2}S_n^2$与$\sqrt{\frac{n}{n+1}}(X_{n+1}-\bar{X})$相互独立

由定义，$\frac{\frac{1}{\sigma}\sqrt{\frac{n}{n+1}}(X_{n+1}-\bar{X})}{\frac{1}{\sigma}\sqrt{\frac{(n-1)}{n-1}}S_n} = \sqrt{\frac{n}{n+1}}(X_{n+1}-\bar{X})/S_n \sim t_{n-1}$

> 6.(P206,22) **设$X_{(1)}, X_{(2)}, \dots, X_{(n)}$为从均匀分布$U(0,1)$中抽取的次序统计量.**
>
> (1)样本量n为多大时，才能使$P(X_{(n)} \geqslant 0.99) \geqslant 0.95$?
>
> (2)求极差$R_n = X_{(n)} - X_{(1)}$的期望.

解：均匀分布有$F(x) = \begin{cases} 0, & x \leqslant 0 \\ x, 0 < x \leqslant 1 & \\ 1, x > 1 & \end{cases}$，$f(x) = I_{(0,1]}$

(1)要求即为$P(\max X_i \geqslant 0.99) \geqslant 0.95$

而$P(\max X_i \geqslant 0.99) = 1 - P(\max X_i < 0.99) = 1 - P(X_i < 0.99, i = 1, 2, \dots, n) = 1 - 0.99^n$

则$P(X_{(n)} \geqslant 0.99) \geqslant 0.95$有$1 - 0.99^n \geqslant 0.95$，解得$n \geqslant 298.07$，$n$至少为299

(2)由题，$0 < x \leqslant 1, F_{X_{(n)}}(x) = P(X_{(n)} \leqslant x) = x^n, f_{X_{(n)}}(x) = nx^{n-1}I_{(0,1]}$
$F_{X_{(1)}}(x) = P(X_{(1)} \leqslant x) = P(X_i\text{中至少有一个小于}x) = \sum_{i=1}^n C_n^i x^i (1-x)^{n-i}$

$$
\begin{aligned}
f_{X_{(1)}}(x) &= \sum_{i=1}^n [\frac{n!}{i!(n-i)!}ix^{i-1}(1-x)^{n-i} - \frac{n!}{i!(n-i)!}x^i(n-i)(1-x)^{n-i-1}] \\
&= \sum_{i=1}^{n-1}[\frac{n!}{(i-1)!(n-i)!}x^{i-1}(1-x)^{n-i} - \frac{n!}{i!(n-i-1)!}x^i(1-x)^{n-i-1}] + nx^{n-1} \\
&= n(1-x)^{n-1}I_{(0,1]}
\end{aligned}
$$

考虑$X_{(1)}$和$X_{(n)}$的联合分布，记U为X_i中不大于u的个数，V为X_i中比u大且不大于v的个数，

$$
\begin{aligned}
F_{X_{(1)},X_{(n)}}(u,v) &= P(X_{(1)} \leqslant u, X_{(n)} \leqslant v) \\
&= P(U \geqslant 1, U+V = n) \\
&= \sum_{k=1}^n P(U = k, V = n-k) \\
&= \sum_{k=1}^n \frac{n!}{k!(n-k)!}F(u)^k(F(v)-F(u))^{n-k}
\end{aligned}
$$

类似$f_{(1)}(x)$有

$$
\begin{aligned}
f_{X_{(1)},X_{(n)}}(u,v) &= \frac{\partial^2}{\partial v \partial u}F_{X_{(1)},X_{(n)}}(u,v) \\
&= \frac{\partial}{\partial v}nf(u)(F(v)-F(u))^{n-1} \\
&= n(n-1)f(u)f(v)(F(v)-F(u))^{n-2}, \quad 0 \leqslant u \leqslant v \leqslant 1
\end{aligned}
$$

则

$$\begin{aligned} f_{R_n}(w) &= \int_{-\infty}^{\infty} f_{X_{(1)},R_n}(x,w)\mathrm{d}x \\ &= \int_{-\infty}^{\infty} f_{X_{(1)},X_{(n)}}(x,x+w)\mathrm{d}x \\ &= \int_0^{1-w} n(n-1)f(x)f(x+w)(F(x+w)-F(x))^{n-2}\mathrm{d}x \\ &= n(n-1)w^{n-2}(1-w)I_{(0,1]} \end{aligned}$$

故

$$ER_n = \int_0^1 wn(n-1)w^{n-2}(1-w)\mathrm{d}w = (n-1)\int_0^1 (1-w)\mathrm{d}w^n = \frac{n-1}{n+1}$$