1 第一次作业

5.29

题目大意：$X_1 \cdots X_n, n = 100$ 为一列的随机变量，均值 $\mu = 1$ 为，标准差 $\sigma = 0.05$，如何估计概率 $P(\sum_i^n X_i > a)$，$a = 100.4$，给出必要的假设.

解答:

方法一： 假设这 100 个随机变量是 iid 的，由中心极限定理，$T = (S - n\mu)/(\sqrt{n}\sigma) = (\sum_i^n X_i - n\mu)/(\sqrt{n}\sigma)$ 近似服从正态分布 $N(0, 1)$.
所以 $P(\sum_i^n X_i > a) = P(\sqrt{n}\sigma T + n\mu > a) = P(T > (a - n\mu)/\sqrt{n}\sigma) = P(T > 0.8)$，查表可得，概率近似值为 0.2119.

方法二（不会考，但是可以了解一下）： 仍然假设这 100 个随机变量是 iid 的，但不用 CLT 近似，假设这些随机变量是" 次高斯"(sub-Gaussain) 的，则由 Hoeffding 不等式，$P(\sum_i^n X_i > a) = P(\sum_i^n X_i - n\mu > a - n\mu) \leq \exp(-\frac{(a-n\mu)^2}{2n\sigma^2}) = 0.73$.
(注：虽然方法二给出的上界要比用正态近似给出的近似值大了许多，但是，但是它给出的是上界是**严格**成立的，因此在实际中应用很多；至于这个不等式的来源，可以参考任意一本高维概率或者高维统计的书，总之它可以看成是 Chebyshev 不等式的进一步发展。)

5.31

题目大意: 背景大致同上，$n = 100, \sigma = 3$，分别用 CLT 与 Chebyshev 多项式给出 $\bar{X} - \mu$ 的置信度为 0.9 的置信区间.

解答:

CLT 方法:$\sqrt{n}(\bar{X} - \mu)/\sigma$ 近似服从 $N(0, 1)$，设 $P(\sqrt{n}|\bar{X} - \mu|/\sigma > a) = 0.1$，则置信区间上为 $[-a\sigma/\sqrt{n}, a\sigma/\sqrt{n}]$，查表知 $a = 1.645$，故 $[-0.4935, 0.4935]$.

Chebyshev: 由 $P(|\bar{X} - \mu| > a) \leq Var(\bar{X})/a^2 = 0.1$，$Var(\bar{X}) = 0.09$, $\Longrightarrow$，置信区间为 $[-0.95, 0.95]$.

5.32

题目大意： 随机变量列 $X_1, \cdots$ 依概率收敛于常数 a，且 $\forall i, P(X_i > 0) = 1$.

(1) 证明:$\sqrt{X_i}$, a/X_i 依概率收敛.
(2)(样本方差 S^2 的相合性) 设 $X_1, X_2, \cdots$ 是一列独立同分布的**连续型**随机变量, 且 $\mathrm{E}X_i = \mu, \mathrm{Var}\, X_i = \sigma^2 < +\infty$. 令

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2.$$

若 $Var(S_n^2) \to 0$, 用 (1) 证明: σ/S_n 依概率收敛于 1.
(3) 事实上，不需要 $Var(S_n^2) \to 0$ 这个条件，仍然可以证明 (2)，请试着证一下.

解答:

(1): 按依概率收敛的定义写:
先证: $a \geq 0$.
事实上, 若 $a < 0$, 取 $\varepsilon = -a/4$, 所以 $P(X_n < 0) \geq P(5a/4 \leq X_n \leq 3a/4) = P(|X_n - a| \leq \varepsilon) \to 1$，与条件矛盾.
若 $a > 0$:
$P(|\sqrt{X_i} - \sqrt{a}| > \varepsilon) = P(|X_i - a| > \varepsilon|\sqrt{X_i} + \sqrt{a}|) \leq P(|X_i - a| > \varepsilon\sqrt{a}) \to 0$.
若 $a = 0$:
$P(|\sqrt{X_i}| > \varepsilon) = P(|X_i| > \varepsilon^2) \to 0$.

对于 a/X_i, 若 $a = 0$,trivial.
若 $a > 0$:
不妨设 $\varepsilon < 1/2$，则 $P(|a/X_i - 1| < \varepsilon) = P(a/(1+\varepsilon) < X_i < a/(1-\varepsilon)) = P(-a\varepsilon/(1+\varepsilon) < X_i - a < a\varepsilon/(1-\varepsilon)) \leq P(-2a\varepsilon < X_i - a < 2a\varepsilon) \to 1$.

事实上，也可以用结论:
X_n 依概率收敛于 X 当且仅当：对于 X_n 任意的子列，存在其子列 a.s. 收敛于 X.
以及: a.s. 收敛套上一个连续函数仍然成立.

(2)
$P(S_n^2 = 0) = P(X_1 = X_2 = \cdots = X_n) = 0$.
又因为 $Var(S_n) \to 0$，故 S_n L^2 收敛于 σ^2. 故 S_n 依概率收敛. 故由 S_n 满足题目 (1) 的条件，先用第二个，再用第一个.

(3)

$$S_n^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}_n\right)^2=\frac{1}{(n-1)}[\sum_{i=1}^{n}X_i^2-\frac{1}{n}(\sum_{i=1}^{n}X_i)^2]=\frac{1}{(n-1)}\sum_{i=1}^{n}X_i^2-\frac{1}{n(n-1)}(\sum_{i=1}^{n}X_i)^2$$

由弱大数律：$\frac{1}{(n-1)}\sum_{i=1}^{n}X_i^2$ 依概率收敛于 EX^2，$\frac{1}{n(n-1)}(\sum_{i=1}^{n}X_i)^2$ 依概率收敛于 μ^2.
所以 S_n^2 一定依概率收敛于 σ^2.
后同 (2).

5.44

题目： 设 $X_i, i=1,2,\cdots$ 是独立的 Bernoulli (p) 随机变量, $Y_n=\frac{1}{n}\sum_{i=1}^{n}X_i$.
(a) 证明 $\sqrt{n}\left(Y_n-p\right)$ 依分布收敛于 $N[0,p(1-p)]$;
(b) 证明: 对任意 $p\neq 1/2$, 方差的估计 $Y_n\left(1-Y_n\right)$ 满足 $\sqrt{n}\left[Y_n\left(1-Y_n\right)-p(1-p)\right]$ 依分布收敛于 $N\left[0,(1-2p)^2p(1-p)\right]$;
(c) 证明: 对于 $p=1/2, n\left[Y_n\left(1-Y_n\right)-\frac{1}{4}\right]$ 依分布收敛于 $-\frac{1}{4}\chi_1^2$.

解答：
(a) 由 CLT 显然.
(b)$g(x)=x(1-x)$，则 $g(x)$ 在 p 处可导，$g'(x)=1-2x$，所以由一阶 delta 定理可证。
(c) 注意此时 $\sigma^2=1/4$，用二阶 delta 定理，$g”(x)=-2$，所以

$$\frac{2n}{\sigma^2 g”(\theta)}[g(T_n)-g(\theta)]=-4n[Y_n(1-Y_n)-\frac{1}{4}]\to\chi^2$$

再由 Slutsky 定理即可.

补充

Delta 方法一些推广：
Prop1: 设 $\sqrt{n}\left(\mathrm{~T_n}-\theta\right)\xrightarrow{\mathrm{d}}\mathrm{N}\left(0,\sigma^2(\theta)\right)$. 令 $\mathrm{g}:\mathbb{R}\mapsto\mathbb{R}$ 在 θ 处 k 阶可微且 $\mathrm{g}^{(\mathrm{j})}(\theta)=0\forall\mathrm{j}<\mathrm{k}$ 但 $\mathrm{g}^{(\mathrm{k})}(\theta)\neq 0$. 则

$$\mathrm{n}^{\mathrm{k}/2}\left\{\mathrm{~g}\left(\mathrm{~T_n}\right)-\mathrm{g}(\theta)\right\}\xrightarrow{\mathrm{d}}\frac{\mathrm{g}^{(\mathrm{k})}(\theta)\sigma^{\mathrm{k}}(\theta)}{\mathrm{k}!}\{\mathrm{N}(0,1)\}^{\mathrm{k}}$$

Prop2： 设 $\sqrt{\mathrm{n}}\left(\boldsymbol{T}_{\mathrm{n}}-\boldsymbol{\theta}\right) \xrightarrow{\mathrm{d}} \mathrm{N}_{\mathrm{p}}(\mathbf{0}, \Sigma(\boldsymbol{\theta}))$. 令 $\boldsymbol{g}: \mathbb{R}^{\mathrm{p}} \mapsto \mathbb{R}^{\mathrm{m}}$ 在 $\boldsymbol{\theta}$ 可微且有非零梯度 $\nabla \mathbf{g}(\boldsymbol{\theta})$(m = 1 时为列向量，m>1 时为 Jacobi 矩阵) 。则

$$\sqrt{\mathrm{n}}\left\{\boldsymbol{g}\left(\boldsymbol{T}_{\mathrm{n}}\right)-\boldsymbol{g}(\boldsymbol{\theta})\right\} \xrightarrow{\mathrm{d}} \mathrm{N}_{\mathrm{m}}\left(\mathbf{0}, \nabla^{\top} \boldsymbol{g}(\boldsymbol{\theta}) \Sigma(\boldsymbol{\theta}) \nabla \boldsymbol{g}(\boldsymbol{\theta})\right)$$

例 1：设 $\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_{\mathrm{n}}$ 为 i.i.d. 随机变量有均值 μ，方差 σ^2 且有 $\mathrm{E}\left(\mathrm{X}_1^4\right)<\infty$ 。则 $\sqrt{\mathrm{n}}\left(\mathrm{S}_{\mathrm{n}}^2-\sigma^2\right) \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, \mu_4-\sigma^4\right)$ 。

证明：

注意到 S_n 的麻烦之处就在于它不是 iid 的和，所以我们就要找 iid 的和来逼近它。

令 $T_n^2=\frac{1}{n} \sum_{i=1}^n\left(X_i-\mu\right)^2$

先证：S_n^2 / T_n^2 依概率收敛于 1.

事实上:

$$\frac{S_n^2}{T_n^2}=\frac{n}{n-1} \frac{\sum_{i=1}^n\left(X_i-\bar{X}\right)^2}{\sum_{i=1}^n\left(X_n-\mu\right)^2}=\frac{n}{n-1} \frac{\sum_{i=1}^n X_i^2-n \bar{X}^2}{\sum_{i=1}^n X_n^2-2 n \bar{X} \mu+n \mu^2}$$

上式中，分子分母同时除以 $\sum_{i=1}^n X_i^2$ 这一项，

注意 $\frac{1}{n} \sum_{i=1}^n X_i^2$ 依概率收敛于 $E X^2$，$\bar{X}$ 依概率收敛于 $E X$.

故由 Slutsky 定理，上述分式依概率收敛于 1.

进一步

$$\frac{\sqrt{n}\left(S_n^2-\sigma^2\right)}{\sqrt{n}\left(T_n^2-\sigma^2\right)}=\frac{S_n^2 / T_n^2-\sigma^2 / T_n^2}{}$$

但是做到这里发现是 0/0 的极限，失败了。

因此，考虑证明：$\sqrt{n}\left(S_n^2-T_n^2\right) \rightarrow 0$（依概率）.

事实上：

$$\begin{aligned}
\sqrt{n}\left(S_n^2-T_n^2\right) & =\sqrt{n}\left[\frac{1}{n-1} \sum_{i=1}^n\left(X_i-\bar{X}\right)^2-\frac{1}{n} \sum_{i=1}^n\left(X_1-\mu\right)^2\right] \\
& =\sqrt{n}\left[\frac{1}{n(n-1)} \sum_{i=1}^n X_i^2-\frac{n}{n-1} \bar{X}^2-\frac{2 \mu}{n} \sum_{i=1}^n X_i-\mu^2\right] \\
& =\sqrt{n}\left[\frac{1}{n(n-1)} \sum_{i=1}^n X_i^2-\frac{1}{n-1} \bar{X}^2-(\bar{X}-\mu)^2\right] \\
& =\frac{1}{\sqrt{n}(n-1)} \sum_{i=1}^n X_i^2-\frac{\sqrt{n}}{n-1} \bar{X}^2-\sqrt{n}(\bar{X}-\mu)^2
\end{aligned}$$

下面，只需要分别证明上面每一项依概率趋于 0:

第一项：由 $\frac{1}{(n-1)} \sum_{i=1}^n X_i^2 \rightarrow E X^2<\infty$ 即可.

第二项：由 $\bar{X}^2 \to \mu^2$ 即可.
第三项：由 $\sqrt{n}(\bar{X}-\mu)$ 依分布收敛于正态，以及 $(\bar{X}-\mu)$ 依概率收敛于 0，再由 Slutsky 定理即可.
至此，我们证明了：$\sqrt{n}(S_n^2 - T_n^2) \to 0$（依概率）.
再用 CLT 可知，$\sqrt{n}(T_n^2-\sigma^2)$ 依分布收敛于正态，且方差恰为 $Var(X^2) = \mu_4 - \sigma^2$.
所以由 Slutsky 定理可知:$\sqrt{n}(S_n^2-\sigma^2)$ 依分布收敛于正态.
另解：
令 $Y_i = (X_i, X_i^2)$，$i = 1 \cdots n$，则由二元的 CLT 定理可知 $\sqrt{n}(\bar{Y}-\vec{\mu}) \to N(0,\Sigma)$，其中 $\vec{\mu} = (EX, EX^2)$.
然后再用二元的 Delta 定理，也可以算出 $\sqrt{n}(S_n^2-\sigma^2)$ 的极限分布.
Remark：
上述两种解法，方法一具有一定的技巧性，需要一点点观察能力，计算量要小点；方法二：多元 CLT+ 多元 DELAT 方法，可以说是典型的套路，适用范围更广，只不过计算量稍微大一点.

令 $\mathrm{g(x)} = \sqrt{\mathrm{x}}$, 由 Delta 方法可知 $\sqrt{\mathrm{n}}\left(\mathrm{S_n}-\sigma\right) \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, \frac{\mu_4-\sigma^4}{4\sigma^2}\right)$.

例 2： 设 $\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X_n}$ 为 i.i.d. 随机变量有均值 μ, 方差 σ^2 且有 $\mathrm{E}\left(\mathrm{X}_1^4\right) < \infty$ 。则有

$$\sqrt{\mathrm{n}}\begin{pmatrix} \overline{\mathrm{X}}_{\mathrm{n}}-\mu \\ \mathrm{S}_{\mathrm{n}}^2-\sigma^2 \end{pmatrix} \xrightarrow{\mathrm{d}} \mathrm{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \mu_3 \\ \mu_3 & \mu_4-\sigma^4 \end{pmatrix}\right)$$

证明: 令 $Y_i = (X_i, X_i^2)$, $i = 1 \cdots n$, 则由二元的 CLT 定理可知 $\sqrt{n}(\bar{Y}-\vec{\mu}) \to N(0,\Sigma)$，其中 $\vec{\mu} = (EX, EX^2)$.
然后再用上面的 Prop2, 将 $\bar{X}$ 与 S_n^2 表示成 $\bar{Y}$ 的映射，然后计算 Jacobi 矩阵，再用 Prop2，即可算出协方差阵.

总结：本周核心内容就是: 大数定律、（多元）CLT、Slutsky、（多元）Delta 方法，以及它们之间组合起来如何灵活运用。