

Ch09: 统计的基本概念

Sampling distribution of Gaussian

December 7, 2023

正态分布的抽样分布

定理 0.67 设 X_1, X_2, \dots, X_n 是来自总体 $\mathcal{N}(\mu, \sigma^2)$ 的样本, 则有

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \sim \mathcal{N}(\mu, \sigma^2/n), \quad \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

定理 0.68 设 X_1, X_2, \dots, X_n 是来自总体 $\mathcal{N}(\mu, \sigma^2)$ 的样本, 其样本均值和无偏样本方差分别为

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

则有

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1).$$

正态分布的抽样分布

定理 0.69 设 X_1, X_2, \dots, X_n 是来自总体 $\mathcal{N}(\mu, \sigma^2)$ 的样本, 其样本均值和无偏样本方差分别为

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

则有 \bar{X} 和 S^2 相互独立, 且

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$

正态分布的抽样分布

定理 0.70 设 X_1, X_2, \dots, X_m 和 Y_1, Y_2, \dots, Y_n 分别来自总体 $\mathcal{N}(\mu_X, \sigma^2)$ 和 $\mathcal{N}(\mu_Y, \sigma^2)$ 的两个独立样本, 令其样本均值分别为 \bar{X} 和 \bar{Y} , 无偏样本方差分别为 S_X^2 和 S_Y^2 , 则有

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}} \sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t(m+n-2).$$

定理 0.71 设 X_1, X_2, \dots, X_m 和 Y_1, Y_2, \dots, Y_n 分别来自总体 $\mathcal{N}(\mu_X, \sigma_X^2)$ 和 $\mathcal{N}(\mu_Y, \sigma_Y^2)$ 的两个独立样本, 令其无偏样本方差分别为 S_X^2 和 S_Y^2 , 则有

$$\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F(m-1, n-1).$$

抽样分布：例 0.124

例 0.124 设随机变量 $T \sim t(n)$, 求 $Y = T^2$ 的分布.

解答：例 0.124

题目：设随机变量 $T \sim t(n)$, 求 $Y = T^2$ 的分布.

解答：

- 根据 t 分布的定义 0.77 可知, 随机变量 $T = \frac{X}{\sqrt{Y/n}}$ 服从自由度为 n 的 t 分布, 其中随机变量 $X \sim \mathcal{N}(0, 1)$ 和 $Y \sim \chi^2(n)$ 相互独立. 因此

$$Y = T^2 = \left(\frac{X}{\sqrt{Y/n}} \right)^2 = \frac{X^2}{Y/n}$$

- 易知 $X^2 \sim \chi^2(1)$, $Y/n \sim \chi^2(1)$, 且 X^2 与 Y/n 相互独立, 因此 $Y = T^2 \sim F(1, n)$.

抽样分布：例 0.125

例 0.125 设 X_1, X_2, \dots, X_5 是来自 $\mathcal{N}(0, 1)$ 的样本, 令 $Y = c_1(X_1 + X_3)^2 + c_2(X_2 + X_4 + X_5)^2$. 求常数 c_1, c_2 使 Y 服从 χ^2 分布.

解答：例 0.125

题目：设 X_1, X_2, \dots, X_5 是来自 $\mathcal{N}(0, 1)$ 的样本，令 $Y = c_1(X_1 + X_3)^2 + c_2(X_2 + X_4 + X_5)^2$. 求常数 c_1, c_2 使 Y 服从 χ^2 分布.

解答：

- 随机变量 $\frac{X_1 + X_3}{\sqrt{2}} \sim \mathcal{N}(0, 1)$, 又 $\frac{X_2 + X_4 + X_5}{\sqrt{3}} \sim \mathcal{N}(0, 1)$ 且相互独立, 根据 χ^2 分布的定义可得

$$\left(\frac{X_1 + X_3}{\sqrt{2}}\right)^2 + \left(\frac{X_2 + X_4 + X_5}{\sqrt{3}}\right)^2 \sim \chi^2(2),$$

- 即 $c_1 = \frac{1}{2}, c_2 = \frac{1}{3}, Y \sim \chi^2(2)$.

抽样分布：例 0.126

例 0.126 设 X_1, X_2 是来自总体 $\mathcal{N}(0, \sigma^2)$ 的样本, 求 $\frac{(X_1+X_2)^2}{(X_1-X_2)^2}$ 的分布.

解答：例 0.126

题目：设 X_1, X_2 是来自总体 $\mathcal{N}(0, \sigma^2)$ 的样本，求 $\frac{(X_1+X_2)^2}{(X_1-X_2)^2}$ 的分布。

解答：

- 随机变量 $\frac{X_1+X_2}{\sqrt{2}\sigma} \sim \mathcal{N}(0, 1)$ ，又 $\frac{X_1-X_2}{\sqrt{2}\sigma} \sim \mathcal{N}(0, 1)$ 且相互独立，根据 χ^2 分布的定义可得

$$\left(\frac{X_1 + X_2}{\sqrt{2}\sigma}\right)^2 \sim \chi^2(1), \quad \left(\frac{X_1 - X_2}{\sqrt{2}\sigma}\right)^2 \sim \chi^2(1),$$

- 根据 F 分布的定义可得

$$\frac{(X_1 + X_2)^2}{(X_1 - X_2)^2} = \frac{\left(\frac{X_1+X_2}{\sqrt{2}\sigma}\right)^2}{\left(\frac{X_1-X_2}{\sqrt{2}\sigma}\right)^2} \sim F(1, 1).$$

抽样分布小结

设 X_1, X_2, \dots, X_m 和 Y_1, Y_2, \dots, Y_n 分别来自总体 $\mathcal{N}(\mu_X, \sigma^2)$ 和 $\mathcal{N}(\mu_Y, \sigma^2)$ 的两个独立样本, 抽样分布如下表所示.

统计量	抽样分布
单个样本均值	$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \sim \mathcal{N}(\mu_X, \sigma^2/n)$
单个样本方差	$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$
两个样本之差	$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2} \sqrt{\frac{1}{m} + \frac{1}{n}}}} \sim t(m+n-2)$
两个样本的方差之比	$\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F(m-1, n-1)$

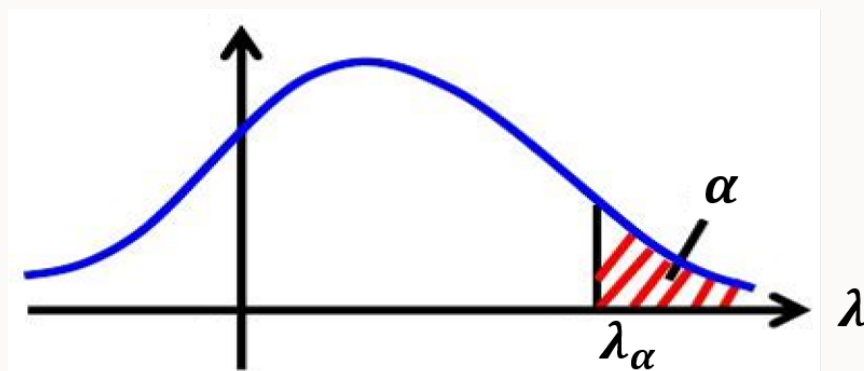
分位数 (点)

很多概率统计问题最后都归结为求解满足概率不等式 $1 - F(x) > \alpha$ 的最大 x , 其解可用上侧分位数 λ_α 表示. 为此人们对常用的分布 (如正态分布、 t 分布、 χ^2 分布等) 编制了各种分位数表供实际使用.

定义 0.79 对给定 $\alpha \in (0, 1)$ 和随机变量 X , 称满足

$$P(X > \lambda_\alpha) = \alpha$$

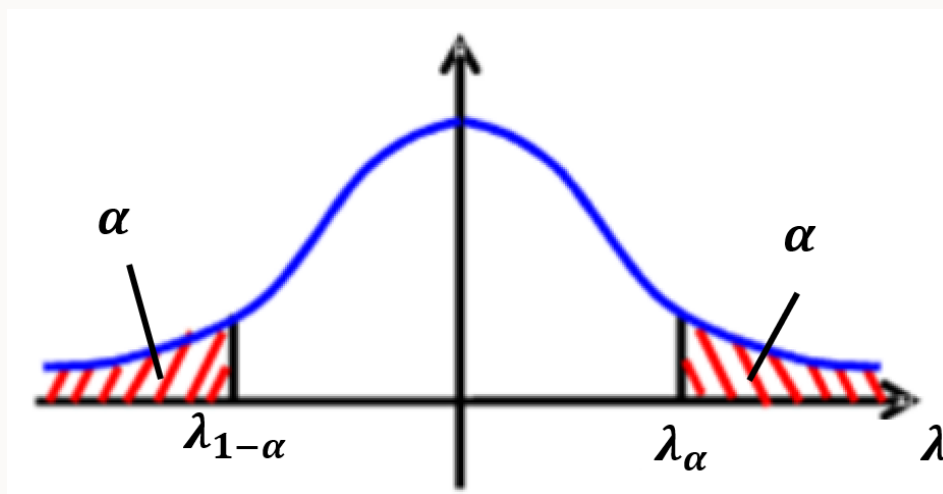
的实数 λ_α 为上侧 α 分位数 (点).



对称分布的分位数

定理 0.72 随机变量 X 的概率密度函数关于 y 轴对称, 则有

$$\lambda_{1-\alpha} = -\lambda_{\alpha}.$$



分位数：例 0.127

例 0.127 设 T_A, T_B 表示某厂生产的两种轴承 A, B 的寿命, 已知 $\lambda_{0.5}^A = 1000\text{h}$, $\lambda_{0.5}^B = 1500\text{h}$, 请说明 A, B 两种轴承中哪个质量更好?

解答：例 0.127

题目：设 T_A, T_B 表示某厂生产的两种轴承 A, B 的寿命, 已知 $\lambda_{0.5}^A = 1000\text{h}$, $\lambda_{0.5}^B = 1500\text{h}$, 请说明 A, B 两种轴承中哪个质量更好?

解答:

- 根据上侧分位数的定义可知, $\lambda_{0.5}^A = 1000\text{h}$ 表示 A 轴承中约有 50% 的寿命超过 1000h, $\lambda_{0.5}^B = 1500\text{h}$ 表示 B 轴承中约有 50% 的寿命超过 1500h, 从上侧 $\alpha = 0.5$ 分位数上说明了后者的质量比前者更高一点.
- 特别的, 称上侧 $\alpha = 0.5$ 的分位数 (点) $\lambda_{0.5}$ 为中位数, 中位数和均值一样都是随机变量的特征数.

正态分布的分位数

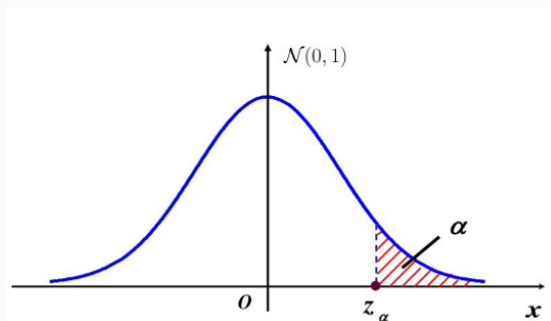
定义 **0.80** 对正态分布 $X \sim \mathcal{N}(0, 1)$, 给定 $\alpha \in (0, 1)$, 满足

$$P(X > \mu_\alpha) = \int_{\mu_\alpha}^{\infty} f(x)dx = \alpha$$

的点 μ_α 称为正态分布上侧 α 分位点.

正态分布的分位数的性质:

- 由对称性可知: $\mu_{1-\alpha} = -\mu_\alpha$
- 由正态分布的密度函数可知: $\Phi(\mu_\alpha) = 1 - \alpha$

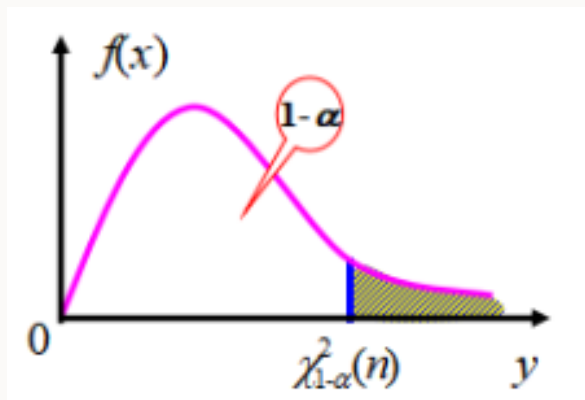


χ^2 分布的分位数

定义 0.81 对 χ^2 分布 $X \sim \chi^2(n)$, 给 $\alpha \in (0, 1)$, 满足 $P(X \geq \chi_\alpha^2(n)) = \alpha$ 的点 $\chi_\alpha^2(n)$ 称为 $\chi^2(n)$ 分布上侧 α 分位点.

χ^2 分布的分位数的性质:

- 当 $n \rightarrow \infty$ 时有 $\chi_\alpha^2(n) \approx \frac{1}{2} (\mu_\alpha + \sqrt{2n-1})^2$, 其中 μ_α 为正态分布上侧 α 分位点.



附录: χ^2 分布的上侧分位数表

附表三 χ^2 分布上侧分位数表 $(P\{\chi^2(n) > \chi^2_\alpha(n)\} = \alpha)$

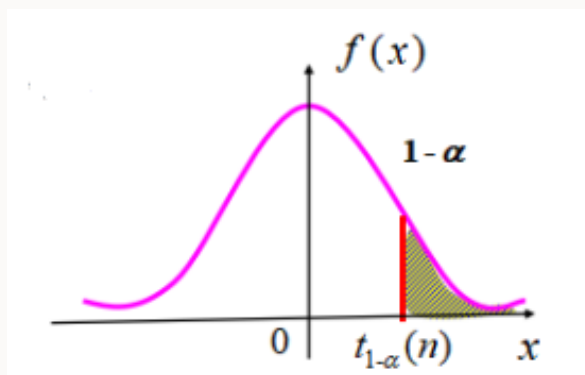
$\alpha \backslash n$	0.995	0.99	0.975	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.025	0.01	0.005
1	0.00004	0.00016	0.001	0.004	0.016	0.102	0.455	1.323	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	0.575	1.386	2.773	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	1.213	2.366	4.108	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	1.923	3.357	5.385	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	2.675	4.351	6.626	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	3.455	5.348	7.841	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	4.255	6.346	9.037	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	5.071	7.344	10.219	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	5.899	8.343	11.389	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	6.737	9.342	12.549	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	7.584	10.341	13.701	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	8.438	11.340	14.845	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	9.299	12.340	15.984	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	10.165	13.339	17.117	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	11.037	14.339	18.245	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	11.912	15.338	19.369	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	12.792	16.338	20.489	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	13.675	17.338	21.605	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	14.562	18.338	22.718	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	15.452	19.337	23.828	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	16.344	20.337	24.935	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	17.240	21.337	26.039	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	18.137	22.337	27.141	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	19.037	23.337	28.241	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	19.939	24.337	29.339	34.382	37.652	40.646	44.314	46.928

t 分布的分位数

定义 0.82 对 t 分布 $X \sim t(n)$, 给 $\alpha \in (0, 1)$, 满足 $P(X \geq t_\alpha(n)) = \alpha$ 的点 $t_\alpha(n)$ 称为 $t(n)$ 分布上侧 α 分位点.

t 分布的分位数的性质:

- 由对称性可知: $t_{1-\alpha} = -t_\alpha$.



附录: t 分布的上侧分位数表

附表四 t 分布上侧分位数表 $(P\{t(n) > t_{\alpha}(n)\} = \alpha)$

$\alpha \backslash n$	0.20	0.15	0.10	0.05	0.025	0.01	0.005
1	1.376	1.963	3.078	6.314	12.706	31.821	63.656
2	1.061	1.386	1.886	2.92	4.303	6.965	9.925
3	0.978	1.25	1.638	2.353	3.182	4.541	5.841
4	0.941	1.19	1.533	2.132	2.776	3.747	4.604
5	0.92	1.156	1.476	2.015	2.571	3.365	4.032
6	0.906	1.134	1.44	1.943	2.447	3.143	3.707
7	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.889	1.108	1.397	1.86	2.306	2.896	3.355
9	0.883	1.1	1.383	1.833	2.262	2.821	3.25
10	0.879	1.093	1.372	1.812	2.228	2.764	3.169
11	0.876	1.088	1.363	1.796	2.201	2.718	3.106
12	0.873	1.083	1.356	1.782	2.179	2.681	3.055
13	0.87	1.079	1.35	1.771	2.16	2.65	3.012
14	0.868	1.076	1.345	1.761	2.145	2.624	2.977
15	0.866	1.074	1.341	1.753	2.131	2.602	2.947
16	0.865	1.071	1.337	1.746	2.12	2.583	2.921
17	0.863	1.069	1.333	1.74	2.11	2.567	2.898
18	0.862	1.067	1.33	1.734	2.101	2.552	2.878
19	0.861	1.066	1.328	1.729	2.093	2.539	2.861
20	0.86	1.064	1.325	1.725	2.086	2.528	2.845

附表四 t 分布上侧分位数表 $(P\{t(n) > t_{\alpha}(n)\} = \alpha)$

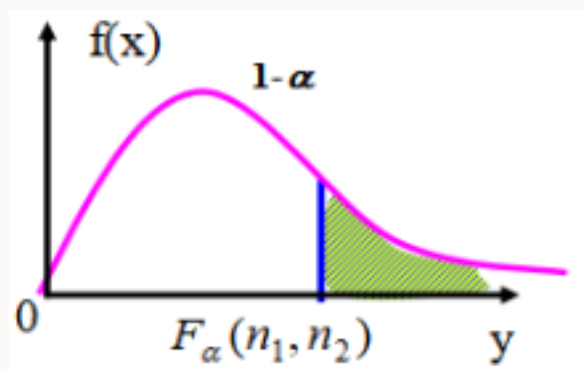
$\alpha \backslash n$	0.20	0.15	0.10	0.05	0.025	0.01	0.005
21	0.859	1.063	1.323	1.721	2.08	2.518	2.831
22	0.858	1.061	1.321	1.717	2.074	2.508	2.819
23	0.858	1.06	1.319	1.714	2.069	2.5	2.807
24	0.857	1.059	1.318	1.711	2.064	2.492	2.797
25	0.856	1.058	1.316	1.708	2.06	2.485	2.787
26	0.856	1.058	1.315	1.706	2.056	2.479	2.779
27	0.855	1.057	1.314	1.703	2.052	2.473	2.771
28	0.855	1.056	1.313	1.701	2.048	2.467	2.763
29	0.854	1.055	1.311	1.699	2.045	2.462	2.756
30	0.854	1.055	1.31	1.697	2.042	2.457	2.75
31	0.8535	1.0541	1.3095	1.6955	2.0395	2.453	2.7441
32	0.8531	1.0536	1.3086	1.6939	2.037	2.449	2.7385
33	0.8527	1.0531	1.3078	1.6924	2.0345	2.445	2.7333
34	0.8524	1.0526	1.307	1.6909	2.0323	2.441	2.7284
35	0.8521	1.0521	1.3062	1.6896	2.0301	2.438	2.7239
36	0.8518	1.0516	1.3055	1.6883	2.0281	2.434	2.7195
37	0.8515	1.0512	1.3049	1.6871	2.0262	2.431	2.7155
38	0.8512	1.0508	1.3042	1.686	2.0244	2.428	2.7116
39	0.851	1.0504	1.3037	1.6849	2.0227	2.426	2.7079
40	0.8507	1.0501	1.303	1.684	2.021	2.423	2.704
60	0.8477	1.0455	1.296	1.671	2.000	2.390	2.660
120	0.8446	1.0409	1.289	1.658	1.98	2.358	2.617
∞	0.8416	1.0364	1.282	1.645	1.96	2.326	2.576

F 分布的分位数

定义 0.83 对 F 分布 $X \sim F(m, n)$, 给 $\alpha \in (0, 1)$, 满足 $P(X \geq F_\alpha(m, n)) = \alpha$ 的点 $F_\alpha(m, n)$ 称为 $F(m, n)$ 分布上侧 α 分位点.

F 分布的分位数的性质:

- $F_{1-\alpha}(m, n) = 1/F_\alpha(n, m)$.



正态分布的抽样分布：例 0.128

例 0.128 设 X_1, X_2, \dots, X_{10} 是来自总体 $\mathcal{N}(\mu, 1/4)$ 的样本. 问

(1) 若 $\mu = 0$, 求 $P(\sum_{i=1}^{10} X_i^2 \geq 4)$

(2) 若 μ 未知, 求 $P(\sum_{i=1}^{10} (X_i - \bar{X})^2 \geq 3.45)$

解答：例 0.128

题目：设 X_1, X_2, \dots, X_{10} 是来自总体 $\mathcal{N}(\mu, 1/4)$ 的样本, i) 若 $\mu = 0$, 求 $P(\sum_{i=1}^{10} X_i^2 \geq 4)$; ii) 若 μ 未知, 求 $P(\sum_{i=1}^{10} (X_i - \bar{X})^2 \geq 3.45)$.

解答：

- 若 $\mu = 0$, 根据 χ^2 分布的定义可知 $\left(\frac{X_1}{1/2}\right)^2 + \left(\frac{X_2}{1/2}\right)^2 + \dots + \left(\frac{X_{10}}{1/2}\right)^2 \sim \chi^2(10)$. 则有 $P\left(\sum_{i=1}^{10} X_i^2 \geq 4\right) = P\left(\frac{\sum_{i=1}^{10} X_i^2}{1/4} \geq \frac{4}{1/4}\right)$. 通过查询 χ^2 分布的上侧分位数表可知, 自由度为 10 的 χ^2 分布上侧 0.1 分位点恰好约为 16, 即 $P(\sum_{i=1}^{10} X_i^2 \geq 4) \approx 0.1$.
- 根据定理 0.67 可知 $\bar{X} = \frac{1}{10} \sum_{i=1}^{10} X_i \sim \mathcal{N}(\mu, 1/40)$, 则有 $\frac{(X_i - \bar{X})}{\sqrt{11/40}} \sim \mathcal{N}(0, 1)$, 根据 χ^2 分布的定义可知 $\frac{\sum_{i=1}^{10} (X_i - \bar{X})^2}{11/40} \sim \chi^2(10)$. 则有

$$P\left(\sum_{i=1}^{10} (X_i - \bar{X})^2 \geq 3.45\right) = P\left(\frac{\sum_{i=1}^{10} (X_i - \bar{X})^2}{11/40} \geq \frac{3.45}{11/40} \approx 12.55\right)$$

通过查询 χ^2 分布的上侧分位数表可知, 自由度为 10 的 χ^2 分布上侧 0.25 分位点恰好约为 12.55, 即 $P(\sum_{i=1}^{10} (X_i - \bar{X})^2 \geq 3.45) \approx 0.25$.

正态分布的抽样分布：例 0.129

例 0.129 设 X_1, X_2, \dots, X_{25} 是来自总体 $\mathcal{N}(12, \sigma^2)$ 的样本, i) 若 $\sigma = 2$, 求 $P(\sum_{i=1}^{25} X_i/25 \geq 12.5)$; ii) 若 σ 未知, 但知道无偏方差为 $S^2 = 5.57$, 求 $P(\sum_{i=1}^{25} X_i/25 \geq 12.95)$.

解答：例 0.129

题目：如上所述.

解答：

- 若 $\sigma = 2$, 根据定理0.67可知 $\bar{X} = \frac{1}{25} \sum_{i=1}^{25} X_i \sim \mathcal{N}(12, 4/25)$, 则有

$$\begin{aligned} P\left(\frac{\sum_{i=1}^{25} X_i}{25} \geq 12.5\right) &= P(\bar{X} \geq 12.5) = P\left(\frac{\bar{X} - 12}{2/5} \geq \frac{12.5 - 12}{2/5}\right) \\ &= 1 - \Phi(1.25) = 0.1056. \end{aligned}$$

- 根据定理 0.68可知 $\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$, 又根据题目有无偏方差为 $S^2 = 5.57$, 即

$$P\left(\frac{\sum_{i=1}^{25} X_i}{25} \geq 12.95\right) = P\left(\frac{\sum_{i=1}^{25} X_i/25 - 12}{5.57/\sqrt{25}} \geq \frac{12.95 - 12}{5.57/\sqrt{25}} \approx 0.85\right)$$

通过查询 t 分布的上侧分位数表可知, 自由度为 24 的 t 分布上侧 0.2 分位点恰好约为 0.85, 即 $P(\sum_{i=1}^{25} X_i/25 \geq 12.95) \approx 0.2$.

如何考察抽样分布？

抽样分布中具有极其复杂的公式, 哪一些知识点是需要我们记忆的, 哪一些是比较重要的?

- 利用中心极限定理, 将变量归一化为 $\mathcal{N}(0, 1)$
 - $Y \sim \chi^2(n)$, which operation of Y obeys $\mathcal{N}(0, 1)$.
 - 设总体分布 $U(0, 1)$, x_1, x_2, \dots, x_n 为样本, 试求 $x_{(k)}$.
- 本节课所提及的六种分布 (包含三大抽样分布), 其来源是怎样的?
 - $X_1, X_2 \sim \mathcal{N}(0, 1)$, which distribution does $X_1^2 + X_2^2$ obey?
 - $X_1, X_2 \sim \mathcal{N}(0, 1)$, if $aX_1^2 + bX_2^2$ obeys χ^2 , find a, b .
 - If $Y \sim t(n)$, compute $Y = T^2$.
- 六种分布 (包含三大抽样分布) 的数字特征 (期望、方差)、基本性质
- 查表 (三大抽样分布、正态分布) 判断分位数及概率
 - 设 X_1, X_2, \dots, X_{10} 是来自总体 $\mathcal{N}(0, 1/4)$ 的样本. 求 $P(\sum_{i=1}^{10} X_i^2 \geq 4)$.