1 Distribuição de Qui-Quadrado

Definição. Uma variável aleatória X é dita possuir distribuição de Qui-Quadrado com parâmetro $k > 0$ se sua fdp é da forma

$$f(x) = \frac{1}{2^{k/2}\Gamma(\frac{k}{2})} x^{k/2-1} e^{-x/2} \underset{(0,\infty)}{I}(x)$$

Notação: $X \sim \chi^2_{(k)}$.

Obs 1.: O parâmetro k é chamado número de graus de liberdade.

1.1 Função Densidade de Probabilidade

Devemos verificar se as condições abaixo são satisfeitas:

1. $f(x) \geq 0$
2. $\int_0^\infty \frac{1}{2^{k/2}\Gamma(\frac{k}{2})} x^{k/2-1} e^{-x/2} dx = 1$

Prova

A primeira condição é óbvia, já que para todo $x > 0$ todos os termos envolvidos são positivos. Para a segunda condição, seja

$$\begin{aligned} I &= \int_0^\infty \frac{1}{2^{k/2}\Gamma(\frac{k}{2})} x^{k/2-1} e^{-x/2} dx \\ &= \frac{1}{2^{k/2}\Gamma(\frac{k}{2})} \int_0^\infty x^{k/2-1} e^{-x/2} dx. \end{aligned}$$

Sabemos que $\int_0^\infty x^{\alpha-1} e^{-x} dx = \Gamma(\alpha)$. Seja $u = \beta x \Rightarrow du = \beta dx,\ \beta > 0$. Assim

$$\int_0^\infty x^{\alpha-1} e^{-\beta x} dx = \frac{\Gamma(\alpha)}{\beta^\alpha} \tag{1}$$

fazendo $\alpha = \frac{k}{2}$ e $\beta = \frac{1}{2}$ temos que

$$\int_0^\infty x^{k/2-1} e^{-x/2} dx = \frac{\Gamma(\frac{k}{2})}{(\frac{1}{2})^{\frac{k}{2}}}$$

logo

$$\left[\frac{1}{2^{k/2}\Gamma(\frac{k}{2})}\right]\left[2^{k/2}\Gamma\left(\frac{k}{2}\right)\right] = 1$$

Obs 2.: Como $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, a fdp de X para $k = 1$ é dada por

$$f(x) = \frac{1}{\sqrt{\pi}2^{1/2}} x^{-\frac{1}{2}} e^{-x/2} = \frac{1}{\sqrt{2\pi x}} e^{-x/2} \underset{(0,\infty)}{I}(x)$$

Obs 3.: A fdp de X para $k = 2$ é dada por

$$f(x) = \frac{1}{2} e^{-x/2} \underset{(0,\infty)}{I}(x)$$

logo $X \sim \chi^2_{(2)} = Exp(\lambda = \frac{1}{2})$.

Obs 4.: O gráfico da fdp de X para $k = 1, 2, 3, 4$ é dado pela **figura 1**. O gráfico da fdp de X para $k \geq 3$ sempre começará na origem.

Obs 5.: O gráfico da fdp de X para $k = 10, 20, 30, 40$ é dado pela **figura 2**.

1.2 Função de Distribuição

A função de distribuição (fd) de $X \sim \chi^2_{(k)}$ é definida por

$$F(x) = \int_0^x \frac{1}{2^{k/2}\Gamma(\frac{k}{2})} t^{k/2-1} e^{-t/2} dt \underset{(0,\infty)}{I}(x)$$

Esta função, para alguns valores de k não têm forma fechada. Ela vem tabelada na maioria dos livros de Estatística.

G.L.	99%	98%	97,5%	95%	90%	80%	70%	50%	30%	20%	10%	5%	4%	2,5%	2%	1%	0.2%	0.1%
1	0.000	0.001	0.001	0.004	0.016	0.064	0.148	0.455	1.074	1.642	2.705	3.841	4.218	5.024	5.412	6.635	9.550	10.828
2	0.020	0.040	0.051	0.103	0.211	0.446	0.713	1.386	2.408	3.219	4.605	5.992	6.438	7.378	7.824	9.210	12.429	13.816
3	0.115	0.185	0.216	0.352	0.584	1.005	1.424	2.366	3.665	4.642	6.251	7.815	8.311	9.348	9.837	11.345	14.796	16.266
4	0.297	0.429	0.484	0.711	1.064	1.649	2.195	3.357	4.878	5.989	7.779	9.488	10.025	11.143	11.668	13.277	16.924	18.467
5	0.554	0.752	0.831	1.145	1.610	2.342	3.000	4.351	6.064	7.289	9.236	11.070	11.644	12.832	13.388	15.086	18.907	20.515
6	0.872	1.134	1.237	1.635	2.204	3.070	3.828	5.348	7.231	8.558	10.645	12.592	13.198	14.449	15.033	16.812	20.791	22.458
7	1.239	1.564	1.690	2.167	2.833	3.822	4.671	6.346	8.383	9.803	12.017	14.067	14.703	16.013	16.622	18.475	22.601	24.322
8	1.647	2.033	2.180	2.733	3.490	4.594	5.527	7.344	9.524	11.030	13.362	15.507	16.171	17.535	18.168	20.090	24.352	26.125
9	2.088	2.532	2.700	3.325	4.168	5.380	6.393	8.343	10.656	12.242	14.684	16.919	17.608	19.023	19.679	21.666	26.056	27.877
10	2.558	3.059	3.247	3.940	4.865	6.179	7.267	9.342	11.781	13.442	15.987	18.307	19.021	20.483	21.161	23.209	27.722	29.588
11	3.054	3.609	3.816	4.575	5.578	6.989	8.148	10.341	12.899	14.631	17.275	19.675	20.412	21.920	22.618	24.725	29.354	31.264
12	3.571	4.178	4.404	5.226	6.304	7.807	9.034	11.340	14.011	15.812	18.549	21.026	21.785	23.337	24.054	26.217	30.957	32.910
13	4.107	4.765	5.009	5.892	7.042	8.634	9.926	12.340	15.119	16.985	19.812	22.362	23.142	24.736	25.471	27.688	32.535	34.528
14	4.660	5.368	5.629	6.571	7.790	9.467	10.822	13.339	16.222	18.151	21.064	23.685	24.485	26.119	26.873	29.141	34.091	36.123
15	5.229	5.985	6.262	7.261	8.547	10.307	11.721	14.339	17.322	19.311	22.307	24.996	25.816	27.488	28.259	30.578	35.628	37.697
16	5.812	6.614	6.908	7.962	9.312	11.152	12.624	15.338	18.418	20.465	23.542	26.296	27.136	28.845	29.633	32.000	37.146	39.252
17	6.408	7.255	7.564	8.672	10.085	12.002	13.531	16.338	19.511	21.615	24.769	27.587	28.445	30.191	30.995	33.409	38.648	40.790
18	7.015	7.906	8.231	9.390	10.865	12.857	14.440	17.338	20.601	22.759	25.989	28.869	29.745	31.526	32.346	34.805	40.136	42.312
19	7.633	8.567	8.906	10.117	11.651	13.716	15.352	18.338	21.689	23.900	27.204	30.143	31.037	32.852	33.687	36.191	41.610	43.820
20	8.260	9.237	9.591	10.851	12.443	14.578	16.266	19.337	22.774	25.038	28.412	31.410	32.321	34.170	35.020	37.566	43.072	45.315
21	8.897	9.915	10.283	11.591	13.240	15.445	17.182	20.337	23.858	26.171	29.615	32.671	33.597	35.479	36.343	38.932	44.522	46.797
22	9.543	10.600	10.982	12.338	14.041	16.314	18.101	21.337	24.939	27.302	30.813	33.924	34.867	36.781	37.660	40.289	45.962	48.268
23	10.196	11.293	11.689	13.091	14.848	17.186	19.021	22.337	26.018	28.429	32.007	35.172	36.131	38.076	38.968	41.638	47.392	49.728
24	10.856	11.992	12.401	13.848	15.659	18.062	19.943	23.337	27.096	29.553	33.196	36.415	37.389	39.364	40.270	42.980	48.812	51.179
25	11.524	12.697	13.120	14.611	16.473	18.940	20.867	24.337	28.172	30.675	34.382	37.653	38.642	40.647	41.566	44.314	50.223	52.620
26	12.198	13.409	13.844	15.379	17.292	19.820	21.792	25.337	29.246	31.795	35.563	38.885	39.889	41.923	42.856	45.642	51.627	54.052
27	12.879	14.125	14.573	16.151	18.114	20.703	22.719	26.336	30.319	32.912	36.741	40.113	41.132	43.194	44.140	46.963	53.023	55.476
28	13.565	14.848	15.308	16.928	18.939	21.588	23.648	27.336	31.391	34.027	37.916	41.337	42.370	44.461	45.419	48.278	54.411	56.892
29	14.257	15.575	16.047	17.708	19.768	22.475	24.577	28.336	32.461	35.139	39.087	42.557	43.604	45.722	46.693	49.588	55.792	58.301
30	14.954	16.306	16.791	18.493	20.599	23.364	25.508	29.336	33.530	36.250	40.256	43.773	44.834	46.979	47.962	50.892	57.167	59.703

1.3 Momentos, assimetria e excesso de curtose

O r-ésimo momento em relação à origem de $X \sim \chi^2_{(k)}$ é dado por

$$\mu_r' = \mathbb{E}(X^r) = \frac{2^r\Gamma(r+\frac{k}{2})}{\Gamma(\frac{k}{2})}.$$

Prova

$$\begin{aligned}
\mathbb{E}(X^r) &= \frac{1}{2^{k/2}\Gamma(\frac{k}{2})}\int_0^{\infty} x^r x^{k/2-1} e^{-x/2} dx \\
&= \left[\frac{2^{r+k/2}\Gamma(r+k/2)}{2^{k/2}\Gamma(\frac{k}{2})}\right] \\
&= \frac{2^r\Gamma(r+k/2)}{\Gamma(k/2)}
\end{aligned}$$

Assim

$$\mathbb{E}(X) = \mu = \frac{2\Gamma(1+k/2)}{\Gamma(k/2)} = \frac{2k}{2} = k$$

$$\mathbb{E}(X^2) = \frac{2^2\Gamma(2+k/2)}{\Gamma(k/2)} = 4\left(1+\frac{k}{2}\right)\frac{k}{2} = 2k\left(1+\frac{k}{2}\right)$$

$$\mathbb{E}(X^3) = \frac{2^3\Gamma(3+k/2)}{\Gamma(k/2)} = 4k\left(2+\frac{k}{2}\right)\left(1+\frac{k}{2}\right)$$

$$\mathbb{E}(X^4) = \frac{2^4\Gamma(4+k/2)}{\Gamma(k/2)} = 8k\left(3+\frac{k}{2}\right)\left(2+\frac{k}{2}\right)\left(1+\frac{k}{2}\right)$$

A variância de $X \sim \chi^2_{(k)}$ é dada por

$$Var(X) = \sigma^2 = 2k$$

Prova

$$\begin{aligned}
Var(X) &= \mathbb{E}(X^2) - \mathbb{E}^2(X) \\
&= 2k\left(1+\frac{k}{2}\right) - k^2 \\
&= 2k + k^2 - k^2 \\
&= 2k
\end{aligned}$$

A assimetria de $X \sim \chi^2_{(k)}$ é dada por

$$\alpha_3 = \sqrt{\frac{8}{k}}$$

Prova

Sabemos que

$$\alpha_3 = \frac{\mu_3}{\sigma^3}$$

em que $\mu_3 = \mathbb{E}[(X-\mu)^3]$. Calculando μ_3 temos

$$\begin{aligned}
\mu_3 &= \mathbb{E}(X^3) - 3\mathbb{E}(X^2)\mu + 3\mathbb{E}(X)\mu^2 - \mu^3 \\
&= 4k\left(2 + \frac{k}{2}\right)\left(1 + \frac{k}{2}\right) - 3k\left[2k\left(1 + \frac{k}{2}\right)\right] + 3k^3 - k^3 \\
&= 4k\left(2 + \frac{3k}{2} + \frac{k^2}{4}\right) - 6k^2 - 3k^3 + 2k^3 \\
&= 8k + 6k^2 + k^3 - 6k^2 - k^3 \\
&= 8k
\end{aligned}$$

logo

$$\begin{aligned}
\alpha_3 &= \frac{8k}{\sqrt{2k}^3} \\
&= \frac{8k}{2k\sqrt{2k}} \\
&= \frac{\sqrt{16}}{\sqrt{2k}} \\
&= \sqrt{\frac{8}{k}}
\end{aligned}$$

O excesso de curtose de $X \sim \chi^2_{(k)}$ é dada por

$$\alpha_4 = \frac{12}{k}$$

Prova

Sabemos que

$$\alpha_4 = \frac{\mu_4}{\sigma^4} - 3$$

em que $\mu_4 = \mathbb{E}[(X-\mu)^4]$. Calculando μ_4 temos

$$\begin{aligned}
\mu_4 &= \mathbb{E}(X^4) - 4\mathbb{E}(X^3)\mu + 6\mathbb{E}(X^2)\mu^2 - 4\mathbb{E}(X)\mu^3 - \mu^4 \\
&= 8k\left(3 + \frac{k}{2}\right)\left(2 + \frac{3k}{2} + \frac{k^2}{4}\right) - \left[16k^2\left(2 + \frac{3k}{2} + \frac{k^2}{4}\right)\right] + 12k^3\left(1 + \frac{k}{2}\right) - 3k^4 \\
&= 48k + 12k^2
\end{aligned}$$

logo

$$\begin{aligned}
\alpha_4 &= \frac{48k + 12k^2}{4k^2} - 3 \\
&= \frac{48k + 12k^2 - 12k^2}{4k^2} \\
&= \frac{48k}{4k^2} \\
&= \frac{12}{k}
\end{aligned}$$

1.4 Moda "M_o"

A moda de $X \sim \chi^2_{(k)}$ é dada por

$$M_o = k - 2,\ k \geq 3$$

para $k = 1$ não existe moda e $k = 2$ a moda é o ponto zero.
Prova

$$f(x) = \frac{1}{2^{k/2}\Gamma(\frac{k}{2})} x^{k/2-1} e^{-x/2}$$

fazendo $g(x) = \ln f(x)$, tem-se:

$$\begin{aligned}
g(x) &= -\frac{k}{2}\ln 2 - \ln\Gamma(k/2) + \left(\frac{k}{2} - 1\right)\ln x - \frac{x}{2} \\
g'(x) &= \frac{k-2}{2x} - \frac{1}{2} \\
g''(x) &= -\frac{k-2}{2x^2} \\
g''(x) &< 0,\ \text{se} \quad k > 2
\end{aligned}$$

Assim, $g'(M_o) = 0$

$$\frac{k-2}{2M_o} = \frac{1}{2}$$

Portanto,

$$M_o = k - 2,\ k > 2$$

Para $k = 2$ tem-se a distribuição exponencial de parâmetro $1/2$, cuja função densidade é sempre descrescente e portanto o máximo ocorre quando $x = 0$. Assim

$$M_o = k - 2,\ k \geq 2$$

1.5 Função Geradora de Momentos

A função geradora de momentos de $X \sim \chi^2_{(k)}$ é definida por

$$M_X(t) = \left(\frac{1}{1-2t}\right)^{k/2}, \; t < \frac{1}{2}$$

Prova

$$\begin{aligned}
M_X(t) &= \mathbb{E}(e^{tX}) = \int_0^\infty e^{tx} \frac{1}{2^{k/2}\Gamma(\frac{k}{2})} x^{k/2-1} e^{-x/2} dx \\
&= \frac{1}{2^{k/2}\Gamma(\frac{k}{2})} \int_0^\infty x^{k/2-1} e^{-x[1/2-t]} dx \quad \text{desde que } \; 1/2 - t > 0 \\
&= \left(\frac{1}{2^{k/2}\Gamma(\frac{k}{2})}\right)\left(\frac{\Gamma(k/2)}{(1/2-t)^{k/2}}\right) \\
&= \left(\frac{1}{1-2t}\right)^{k/2}, \; t < \frac{1}{2}
\end{aligned}$$

1.6 Função geradora de cumulantes

A função geradora de momentos de $X \sim \chi^2_{(k)}$ é definida por

$$K_X(t) = -\frac{k}{2} \ln(1-2t), \; t < \frac{1}{2} \tag{2}$$

Prova

Sabemos que

$$K_X(t) = \ln M_X(t)$$

assim

$$\begin{aligned}
K_X(t) &= \ln(1-2t)^{-k/2} \\
&= -\frac{k}{2} \ln(1-2t).
\end{aligned}$$

Vamos encontrar o valor esperado e variância a partir dessa função. Derivando a função (2) em relação à t obtemos

$$K'_X(t) = -\frac{k}{2} \cdot \frac{-2}{1-2t} = \frac{k}{1-2t}, \; t < \frac{1}{2}$$

sabemos que $K'_X(0) = \mathbb{E}(X)$. Assim

$$\mathbb{E}(X) = K'_X(0) = \frac{k}{1-0} = k$$

Conhecemos que $K''_X(0) = \text{Var}(X)$. Assim

$$K''_X(t) = \frac{2k}{(1-2t)^2}, \; t < \frac{1}{2}$$

e

$$\text{Var}(X) = K''_X(0) = \frac{2k}{(1-0)^2} = 2k$$

1.7 Distribuição Qui-quadrado no R

No software gratuito conhecido como R, disponível para download no endereço *www.r-project.org*, temos que, em cada distribuição de probabilidade é possível usar algumas operações representadas por uma única letra em seguida a simbologia da distribuição estudada, dentre essas operações destacamos:

`d` calcula a densidade de probabilidade $f(x)$ no ponto. O primeiro termo é o valor que x assume da densidade, o segundo é o número de graus de liberdade

```
dchisq(x, df, ncp=0, log = FALSE)
```

`p` calcula a função de probabilidade acumulada $F(x) = P(X < x)$ no ponto. O primeiro termo é o valor de x que representa o quantil e o terceiro, se for dado como verdade, argumento calcula $P(X \leq x)$, caso contrário, $P(X > x)$;

```
pchisq(q, df, ncp=0, lower.tail = TRUE)
```

`q` calcula o quantil de uma distribuição em relação a uma dada probabilidade, ou seja, calcula o valor de c em que $P(X < c) = p$. É a operação inversa da proposta acima, em que o primeiro termo indica a probabilidade que a função assume;

```
qchisq(p, df, ncp=0, lower.tail = TRUE)
```

`r` gera amostra da distribuição. O primeiro argumento representa o número de observações na amostra gerada.

```
rchisq(n, df, ncp=0)
```

Obs 14: Há outros argumentos dessas funções, tais como `ncp` que indica o parâmetro de não-centralidade e por `log` ou `log.p` se utilizado como verdade substitui o valor das probabilidade p por $\ln p$, resolvemos omitir porque não objeto do nosso estudo.

1.7.1 Comandos básicos

Para mais informações, e exemplos basta utilizar a função `help`.

```
> help(dchisq)
```

Vamos ver alguns exemplos.

```
> dchisq(4,df=3)
[1] 0.1079819
```

que de forma análoga

```
> fx=1/(2^(3/2)*gamma(3/2))*(4)^((3/2)-1)*exp(-4/2);fx
[1] 0.1079819

> pchisq(4,3)
[1] 0.7385359

> qchisq(0.95,12)
[1] 21.02607
```

```
> rchisq(10,1)
 [1] 0.016289145 0.574638864 0.034978639 0.211442672 2.149133465 0.031807597
 [7] 1.408106132 0.009895398 0.910755996 0.605715045
```

Obs 10.: Notamos que a cada geração de uma amostra aleatória obtemos valores diferentes, pois o R muda a "semestre"da geração. Se quisermos uma amostra igual em todas as gerações devemos utilizar o comando `seed(.)`por exemplo

```
> set.seed(945)
> rchisq(7,4)
[1] 5.298402 4.248953 2.296971 2.783382 9.483237 1.903036 8.307755

> rchisq(7,4)
[1] 2.133257 4.228787 5.189371 7.715549 2.464150 3.985626 1.139770

> set.seed(945)
> rchisq(7,4)
[1] 5.298402 4.248953 2.296971 2.783382 9.483237 1.903036 8.307755
```

Vamos agora ver alguns gráficos gerados pelo R a partir do comando `plot` .Podemos gerar gráficos tanto da função densidade $f(x)$ como da função acumulada $F(x)$. Abaixo usamos a distribuição com 6 graus de liberdade

```
plot(function(x) dchisq(x, 6), 0,20)
plot(function(x) pchisq(x, 6), 0,20)
```

Faça como exercício o gráfico com $k = 1, 2, 3$ e comente o resultado.

Os gráficos do R possuem várias opções, como mudar o título do gráfico e dos eixos. Podemos também definir os limites dos eixos, colocar legendas, alterar tamanho da fonte, entre outras opções (ver figura 4)

Figura 4: Gráfico da função densidade de $X \sim \chi^2_{(15)}$ com área hachurada representando $P(X \leq 12)$

O comando para a figura 4 é dado por

```
x=seq(0,36,l=150)
fx=dchisq(x,15)
plot(x,fx,type="l", main="Distribuição Qui-quadrado com 15 g.l.",ylab="f(x)")
ax=c(0,0,x[x<12],12,12)
ay=c(0,dchisq(c(0,x[x<12],12),15),0)
polygon(ax,ay,dens=10)
```

1.7.2 Relações entre distribuições de probabilidade

Um recurso importante no R é construir várias simulações para fazer inferências sobre as distribuições amostrais nas características que temos interesse. A utilização das simulações é importante quando desejamos obter resultados teóricos que não são conhecidos. Um dos resultados que podemos explanar é a aproximação entre a distribuição normal e a qui-quadrado visto nas secções anteriores. Vamos mostrar para o caso, se $X \sim N(0,1)$ então $Y = X^2 \sim \chi^2_{(1)}$.

```
x=rnorm(10000)
hist(x,prob=T)
curve(dnorm(x),add=T)
```

usamos o argumento `prob=T` assim obtemos o valor das frequências relativa, podendo assim comparar com as curva teórica. Fazendo agora o gráfico da qui-quadrado.

```
hist(x^2,prob=T)
curve(dchisq(x,1),0,10,add=T)
```

Na **figura 5** temos os histogramas a partir da amostra gerada dos gráficos da distribuição normal padrão e da distribuição qui-quadrado com $k = 1$.

Figura 5

1.7.3 Distribuição amostral

Vamos ilustrar uma simulação na qual $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$, com $S^2 = \sum_{i=1}^{n} \frac{x_i - \bar{x}}{n-1}$ sendo a variância amostral. Vamos executar os seguintes passos

i) Escolha os parâmetros de uma distribuição normal;

ii) Escolha o tamanho da amostra n e á quantidade de simulações N;

iii) Agora simule N amostras de tamanho n;

iv) Para cada valor da amostra calcule $V = \frac{(n-1)S^2}{\sigma^2}$

v) reproduza um histograma com os valores de V e compare com a curva da distribuição $\chi^2_{(n-1)}$.

Um exemplo para a simulação acima é

```
n=20
N=1000

y=matrix(rnorm(n*N,100,6),nc=N);y
S2=apply(y,2, function(x) {(n-1)*sum((x-mean(x))^2)/(n-1)/36});S2
mean(S2)
var(S2)
hist(S2,prob=T,main="Histograma de V",ylab="f(v)",xlab="v")
curve(dchisq(x,n-1),0,50,add=T)
```

1.8 Transformações importantes

A distribuição qui-quadrado têm várias aplicações em inferência estatística tanto em teste de hipóteses e intervalos de confiança quanto em estimação de variâncias. Um exemplo são os problemas envolvendo a análise de variância, através da distribuição F de Snedecor que é a razão entre duas distribuições qui-quadrados independentes divididas por seus respectivos graus de liberdade. Vamos tratar de algumas relações de transformações de variáveis

Se $X_1, X_2, \cdots, X_n$ são variáveis aleatórias independentes e identicamente distribuídas com distrbuição Normal com média μ_i e variância σ_i^2, então

$$V = \sum_{i=1}^{k} \left(\frac{X_i - \mu_i}{\sigma_i} \right)^2 \tag{3}$$

têm distribuição Qui-Quadrado com k graus de liberdade.

Obs 6.: Uma relação importante é se $X \sim \chi^2_{(1)}$ e $Y = \sum_{i=1}^{k} X_i$ obteremos que $Y \sim \chi^2_{(k)}$.

Obs 7.: É importante ressaltar que se $X_1, X_2, \cdots, X_n \sim N(\mu, \sigma^2)$ e fazendo $V = \sum_{i=1}^{k} \left(\frac{X_i - \mu}{\sigma}\right)^2$, isto é, as médias e variâncias são iguais em todas as variáveis aleatórias, também obteremos que $V \sim \chi^2_{(k)}$.

Obs 8.: Se $X \sim N(\mu_i, 1)$ com médias diferentes de zero temos que se $V = \sum_{i=1}^{k} X_i^2$ então V tem distribuição qui-quadrado não-central.

Teorema 1. Se $Z_1, Z_2, \cdots, Z_k$ é uma amostra aleatória com distribuição normal padrão, então:

a) $\bar{Z} \sim N(0, 1/k)$;

b) $\bar{Z}$ e $\sum_{i=1}^{k} (Z_i - \bar{Z})^2$ são independentes;

c) $\sum_{i=1}^{k} (Z_i - \bar{Z})^2 \sim \chi^2_{(k)}$.

Corolário. Se $X \sim N(\mu, \sigma^2)$ e $X_1, \cdots, X_k$ é uma amostra aleatória de X, então $S^2 = \frac{\sum_{i=1}^{k} (X_i - \bar{X})^2}{k-1}$ é chamada de variância amostral. Com base no teorema concluímos que

$$\frac{(k-1)S^2}{\sigma^2} \sim \chi^2_{(k-1)}$$

Teorema 2. Vamos supor que $X \sim \chi^2_{(k)}$ e se k for suficientemente grande temos que $U = \sqrt{2X}$ tem aproximadamente distribuição normal com média $\sqrt{2k-1}$ e variância 1.

Obs 9: Uma outra aproximação importante é se $X \sim \chi^2_{(k)}$ então $\sqrt[3]{X/k}$ tem aproximadamente distribuição normal com média $1 - \frac{2}{9k}$ e variância $\frac{2}{9k}$.

Obs 10: Seja $X_1 \sim \chi^2_{k_1}$ e $X_2 \sim \chi^2_{k_2}$ variáveis aleatórias independentes. Fazendo $Y = \frac{X_1/k_1}{X_2/k_2}$ temos que $Y \sim F_{(k_1, k_2)}$

Obs 11: Seja $\mathbf{Z}^\top = (Z_1, Z_2, \cdots, Z_n)$, em que os Z_i's são variáveis aleatórias independentes e $Z_i \sim N(0, \sigma^2)$ ou $\mathbf{Z} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I})$, em que a média $\mathbf{0}$ é uma matriz de zeros de ordem n, $\mathbf{I}$ é a matriz indentidade de ordem n e seja $\mathbf{A}$ é uma matriz quadrada $n \times n$ idempotente e com posto $n-k$ e então a forma quadrática $\dfrac{\mathbf{z}^\top\mathbf{A}\mathbf{z}}{\sigma^2} \sim \chi^2_{(n-k)}$

Obs 12: Seja $F(x|k)$ a função de distribuição de $X \sim \chi^2_{(k)}$, então:

a) $F(x|k) = \frac{1}{\Gamma(k/2)} \sum_{i=0}^{\infty} \frac{(-1)^i (x/2)^{k/2+i}}{i!\Gamma(k/2+i)}$;

b) $F(x|k+2) = F(x|k) - \frac{(x/2)^{k/2} e^{-x/2}}{\Gamma(k/2+1)}$;

c) $F(x|2k) = 1 - 2\sum_{i=1}^{k} f(x|2i)$;

d) $F(x|2k+1) = 2\Phi(\sqrt{x}) - 1 - 2\sum_{i=1}^{k} f(x|2i+1)$.

em que $f(x|k)$ é a função densidade de probabilidade de $X \sim \chi^2_{(k)}$ e Φ representa a função acumulada da normal padrão.

Obs 13: Indentidade de Haff: Seja f e h valores reais de uma função, e seja $X \sim \chi^2_{(k)}$, então

$$\mathbb{E}[f(X)h(X)] = \mathbb{E}\left[f(X)\frac{\partial h(X)}{\partial X}\right] + \mathbb{E}\left[\frac{\partial f(X)}{\partial X}h(X)\right] + (n-2)\mathbb{E}\left[\frac{f(X)h(X)}{X}\right]$$

1.9 Geração de números aleatórios

Como visto nas subseções anteriores temos que a soma de k variáveis aleatórias com distribuição normal padrão ao quadrado resulta em uma distribuição qui-quadrado com k graus de liberdade. Podemos usar esse resultado na construção de técnicas de geração de números pseudoaleatórios da distribuição qui-quadrado. Seja a transformação na qual são dados dois números pseudoaleatórios, γ_1 e γ_2, com distribuição uniforme padrão a partir da transformação

$$z_1 = \sqrt{-2\ln\gamma_1}\cos 2\pi\gamma_2$$

e

$$z_2 = \sqrt{-2\ln\gamma_1}\text{sen}2\pi\gamma_2$$

em que z_1 e z_2, são dois números pseudoaleatórios com distribuição normal padrão. Adicionando k números pseudoaleatórios temos que

$$x_{2k} = -2\ln(\gamma_1, \cdots, \gamma_k)$$

e

$$z_{2k+1} = -2\ln(\gamma_1, \cdots, \gamma_k) - 2\ln\gamma_{k+1}\cos^2 2\pi\gamma_{k+2}$$

para k um número inteiro positivo teremos uam distribuição de qui-quadrado com grau de liberdade par ou ímpar. Com esse procedimento muitas operações são evitadas. Também sabemos que a distribuição qui-quadrado é um caso particular da distribuição Gama e podemos conseguir geração de números aleatórios a partir dessa distribuição.

1.10 Intervalos de confiança para a variância

Se $X_1, \cdots, X_n$ são variáveis aleatórias de $X \sim N(\mu, \sigma^2)$ e vimos que $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$. Um intervalo de confiança para a variância com nível de confiança $1-\alpha$ é dado por

$$\frac{(n-1)s^2}{\chi^2_{(1-\alpha/2,n-1)}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{(\alpha/2,n-1)}}$$

em que $\chi^2_{(\alpha;n)}$ o quantil de ordem $1-\alpha$ da distribuição qui-quadrado com n graus de liberdade.

1.11 Testes de hipóteses

Seja uma amostra aleatória, $X_1, \cdots, X_n$ de $X \sim N(\mu.\sigma^2)$. Para testar a hipótese $H_0 : \sigma^2 = \sigma_0^2$ contra $H_1 : \sigma^2 \neq \sigma_0^2$ com nível de significância α, rejeitaremos H_0 se $\dfrac{(n-1)s^2}{\sigma_0^2} \sim \chi^2_{(n-1)}$ for menor que $\chi^2_{\alpha/2,n-1}$ ou maior que $\chi^2_{1-\alpha/2,n-1}$.

1.12 Exercícios

1. Seja $\chi^2_{(\alpha;k)}$ representando a expressão $P(\chi^2_{(k)} > \chi^2_{(\alpha;k)}) = \alpha$, em que $\chi^2_{(\alpha;k)}$ é quantil de ordem $(1-\alpha)$ da distribuição $\chi^2_{(k)}$. Calcule as probabilidades usando a tabela e compare os valores no R

a) $\chi^2_{(0,05;16)}$;

b) $\chi^2_{(0,01;7)}$;

c) $\chi^2_{(0,975;22)}$;

d) $1 - \chi^2_{(0,8;1)}$;

e) $P(\chi^2_{(23)} > 20)$;

f) $P(\chi^2_{(11)} \leq 20,412)$;

g) $P(9,39 < \chi^2_{(18)} < 32,346)$.

h) Ache o valor de a tal que $P(4 < \chi^2_{(11)} < a) = 0,9$.

2. Calcule a distribuição de S^2.

3. Demosntre a igualdade da equação (3).

4. Seja as variáveis aleatórias X, Y e Z com distribuições normais padrão. Gere valores aleatórios no R e mostre, graficamente, que a variável $U = X^2 + Y^2 + Z^2$ tem distribuição Qui-Quadrado(3).