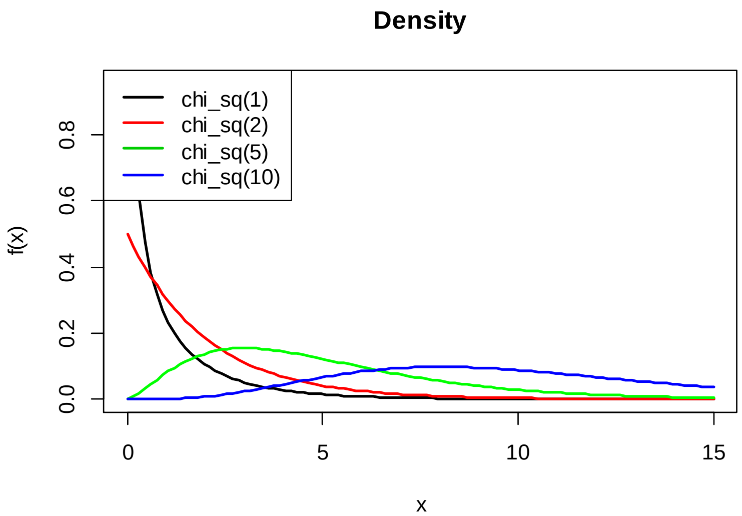
Model rozkładu normalnego

Zakładamy, że badana cecha X (obserwowana zmienna X) ma rozkład normalny. To znaczy, że losowa próba prosta $\boldsymbol{X} = (X_1, \dots, X_n)$ jest ciągiem niezależnych zmiennych losowych o rozkładzie normalnym $N(\mu, \sigma)$.

Wtedy statystyka zwana średnią z próby $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ ma rozkład normalny $N(\mu, \frac{\sigma}{\sqrt{n}})$ zakładając, że populacja, z której wylosowano próbę jest nieskończona i przeliczalna. Statystyka $\bar{X}$ jest nieobciążonym estymatorem średniej populacyjnej μ ponieważ $E(\bar{X}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = E(X_1) = \mu$.

Jeśli wariancja populacyjna σ^2 jest nieznana (i średnia populacyjna również μ) wybieramy jeden z dwóch estymatorów $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ and $\tilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Pierwszy z nich jest nieobciążony dla parametru σ^2 i ma minimalną wariancję, drugi jest obciążony, ale pochodzi z metody największej wiarogodności.

Statystyka $\frac{(n-1)S^2}{\sigma^2} = \frac{n\tilde{S}^2}{\sigma^2}$ ma rozkład chi-kwadrat z $(n-1)$ stopni swobody.



Gęstość rozkładu chi-kwadrat z k stopniami swobody

$$f(x) = \frac{0.5^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}} I_{(0,\infty)}(x)$$

Wartość oczekiwana: k

Wariancja: $2k$

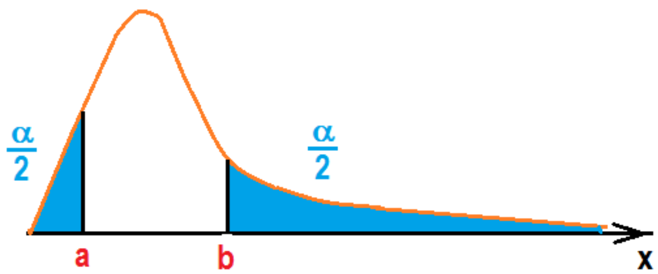
Statystyka $\frac{(n-1)S^2}{\sigma^2} = \frac{n\tilde{S}^2}{\sigma^2}$ może być użyta jako funkcja centralna do konstrukcji przedziału ufności dla parametru σ^2. Przedział ma postać

$$\left(\frac{(n-1)S^2}{\chi^2(1-\frac{\alpha}{2}, n-1)}, \frac{(n-1)S^2}{\chi^2(\frac{\alpha}{2}, n-1)}\right)$$

gdzie $\chi^2(p, n-1)$ jest kwantylem rzędu p z rozkładu χ^2 z $(n-1)$ stopniami swobody.

Przykładowa konstrukcja przedziału ufności:

1. Funkcja centralna $\frac{(n-1)S^2}{\sigma^2}$ ma rozkład $\chi^2(n-1)$ i zależy od parametru σ^2 (ale jej rozkład nie zależy).
2. Ustalamy poziom ufności $1-\alpha$ jako liczbę nieco mniejszą od 1.
3. Zapisujemy wzór (dla przedziału dwustronnego) $P\left(a < \frac{(n-1)S^2}{\sigma^2} < b\right) = 1-\alpha$,
   dla przedziału lewostronnego $P\left(c < \frac{(n-1)S^2}{\sigma^2}\right) = 1-\alpha$, $(c > 0)$
   dla przedziału prawostronnego $P\left(\frac{(n-1)S^2}{\sigma^2} < d\right) = 1-\alpha$, $(d > 0)$.
4. Przekształcamy nierówność w nawiasie wyznaczając parametr.
   $P\left(\frac{(n-1)S^2}{b} < \sigma^2 < \frac{(n-1)S^2}{a}\right) = 1-\alpha$,
   $P\left(\sigma^2 < \frac{(n-1)S^2}{c}\right) = 1-\alpha$,
   $P\left(\sigma^2 > \frac{(n-1)S^2}{d}\right) = 1-\alpha$.
5. Korzystając z wykresu gęstości odczytujemy wartości kwantyli
   $a = \chi^2(\frac{\alpha}{2}, n-1)$, $b = \chi^2(1-\frac{\alpha}{2}, n-1)$, $c = \chi^2(\alpha, n-1)$, $d = \chi^2(1-\alpha, n-1)$.



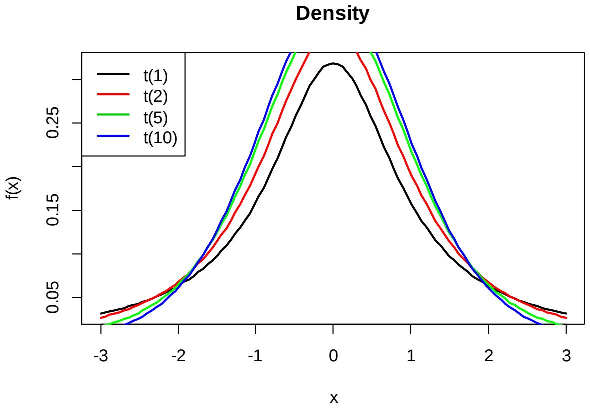
6. Zapisujemy ostateczną postać przedziału
   $P\left(\frac{(n-1)S^2}{\chi^2(1-\frac{\alpha}{2}, n-1)} < \sigma^2 < \frac{(n-1)S^2}{\chi^2(\frac{\alpha}{2}, n-1)}\right) = 1-\alpha$
   $P\left(\sigma^2 < \frac{(n-1)S^2}{\chi^2(\alpha, n-1)}\right) = 1-\alpha$,
   $P\left(\sigma^2 > \frac{(n-1)S^2}{\chi^2(1-\alpha, n-1)}\right) = 1-\alpha$.
7. Jeśli poszukujemy przedziału ufności np. dla odwrotności lub innej funkcji parametru, to nierówność w nawiasie trzeba odpowiednio przekształcić.
   $P\left(\frac{n\cdot\chi^2(\frac{\alpha}{2}, n-1)}{(n-1)S^2} < \frac{n}{\sigma^2} < \frac{n\cdot\chi^2(1-\frac{\alpha}{2}, n-1)}{(n-1)S^2}\right) = 1-\alpha$, itp.

……………

Jeżeli μ oraz σ^2 są nieznane, przedział ufności dla parametru μ ma postać

$$\left(\bar{X} - \frac{S}{\sqrt{n}} t\left(1-\frac{\alpha}{2}, n-1\right), \bar{X} + \frac{S}{\sqrt{n}} t\left(1-\frac{\alpha}{2}, n-1\right)\right)$$

gdzie $t(p, n-1)$ jest kwantylem rzędu p z rozkładu t-Studenta z $(n-1)$ stopniami swobody.



Gęstość rozkładu t-Studenta z k stopniami swobody

$$f(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\sqrt{n\pi}}\left(1+\frac{t^2}{k}\right)^{-\frac{k+1}{2}}$$

Wartość oczekiwana: 0 for $k > 1$

Wariancja: $\frac{k}{k-2}$ for $k > 2$.

Uwaga. Jeśli nie ma statystyki, która nadawałaby się na funkcję centralną, można skorzystać z centralnego twierdzenia granicznego. Wtedy musimy bazować na dużych próbach. Wówczas parametr *exact* = F w poleceniu *enorm* w *EnvStats*.

Rozkładem granicznym statystyki $\bar{X}$ jest $N(\mu, \frac{\sigma}{\sqrt{n}})$, więc jako funkcję centralną można wziąć $\frac{\bar{X}-\mu}{S}\sqrt{n}$ o rozkładzie $N(0,1)$. Wtedy dwustronny przedział ufności ma postać

$$\left(\bar{X} - \frac{S}{\sqrt{n}} z\left(1-\frac{\alpha}{2}\right), \bar{X} + \frac{S}{\sqrt{n}} z\left(1-\frac{\alpha}{2}\right)\right)$$

gdzie $z(p)$ jest kwantylem rzędu p z rozkładu $N(0,1)$.

Model rozkładu wykładniczego

Zakładamy, że badana cecha X (obserwowana zmienna X) ma rozkład wykładniczy $Ex(\lambda)$. To znaczy, że losowa próba prosta $\boldsymbol{X} = (X_1, \ldots, X_n)$ jest ciągiem niezależnych zmiennych losowych o rozkładzie normalnym $Ex(\lambda)$.

Wtedy statystyka postaci $\frac{1}{\bar{X}} = \frac{n}{\sum_{i=1}^{n} X_i}$ jest estymatorem największej wiarogodności parametru λ, a statystyka $\frac{n-1}{n\bar{X}}$ jest estymatorem nieobciążonym o minimalnej wariancji parametru λ.

Jako funkcję centralną dla parametru λ wybiera się statystykę $2n\lambda\bar{X}$ o rozkładzie $\chi^2(2n)$. Na jej podstawie konstruuje się przedział ufności dla parametru λ

$$P\left(\frac{\chi^2(\frac{\alpha}{2}, 2n)}{2n\bar{X}} < \lambda < \frac{\chi^2(1-\frac{\alpha}{2}, 2n)}{2n\bar{X}}\right) = 1-\alpha.$$

Jeśli zmienna losowa X ma rozkład wykładniczy $Ex(\lambda)$, to $E(X) = \frac{1}{\lambda}$ i $Var(X) = \frac{1}{\lambda^2}$. Dlatego przedziały ufności dla wartości oczekiwanej i wariancji w populacji wykładniczej mają postać

$$P\left(\frac{2n\bar{X}}{\chi^2(1-\frac{\alpha}{2}, 2n)} < \frac{1}{\lambda} < \frac{2n\bar{X}}{\chi^2(\frac{\alpha}{2}, 2n)}\right) = 1 - \alpha$$

oraz

$$P\left(\frac{(2n\bar{X})^2}{\left(\chi^2(1-\frac{\alpha}{2}, 2n)\right)^2} < \frac{1}{\lambda^2} < \frac{(2n\bar{X})^2}{\left(\chi^2(\frac{\alpha}{2}, 2n)\right)^2}\right) = 1 - \alpha.$$