Statistika

Vanja Wagner

3. Osnove vjerojatnosti

Aproksimacija binomne distribucije normalnom - nastavak

Kako koristimo ranije spomenuti rezultat gdje za velike n ($np \geqslant 5$ i $nq \geqslant 5$) distribuciju binomne slučajne varijable $X \sim B(n, p)$ možemo aproksimirati normalnom distribucijom,

$$X \approx N(np, npq)?$$

Aproksimacija binomne distribucije normalnom - nastavak

Kako koristimo ranije spomenuti rezultat gdje za velike n ($np \geqslant 5$ i $nq \geqslant 5$) distribuciju binomne slučajne varijable $X \sim B(n,p)$ možemo aproksimirati normalnom distribucijom,

$$X \approx N(np, npq)?$$

Neka je $F = F_{N(np,npq)}$ funkcija distribucije normalne razdiobe $N(np, npq)$. Tada

$$\mathbf{P}(X \leqslant b) \approx F(b) \text{ ili } F(b + \tfrac{1}{2}),$$
$$\mathbf{P}(a \leqslant X \leqslant b) \approx F(b) - F(a) \text{ ili } F(b + \tfrac{1}{2}) - F(a + \tfrac{1}{2}),$$

gdje je crveni izraz *bez korekcije* a plavi izraz *s korekcijom* po neprekidnosti. Obe aproksimacije su korisne, korekcija daje nešto bolju aproksimaciju uzimajući u obzir činjenicu da je binomna razdioba diskretna a normalna razdioba neprekidna.

Aproksimacija binomne distribucije normalnom - nastavak

Kako koristimo ranije spomenuti rezultat gdje za velike n ($np \geqslant 5$ i $nq \geqslant 5$) distribuciju binomne slučajne varijable $X \sim B(n, p)$ možemo aproksimirati normalnom distribucijom,

$$X \approx N(np, npq)?$$

Neka je $F = F_{N(np,npq)}$ funkcija distribucije normalne razdiobe $N(np, npq)$. Tada

$$\begin{aligned} \mathbf{P}(X \leqslant b) &\approx F(b) \text{ ili } F(b + \tfrac{1}{2}), \\ \mathbf{P}(a \leqslant X \leqslant b) &\approx F(b) - F(a) \text{ ili } F(b + \tfrac{1}{2}) - F(a + \tfrac{1}{2}), \end{aligned}$$

gdje je crveni izraz *bez korekcije* a plavi izraz *s korekcijom* po neprekidnosti.

Aproksimacija binomne distribucije normalnom - nastavak

Kako koristimo ranije spomenuti rezultat gdje za velike n ($np \geqslant 5$ i $nq \geqslant 5$) distribuciju binomne slučajne varijable $X \sim B(n, p)$ možemo aproksimirati normalnom distribucijom,

$$X \approx N(np, npq)?$$

Neka je $F = F_{N(np,npq)}$ funkcija distribucije normalne razdiobe $N(np, npq)$. Tada

$$\mathbf{P}(X \leqslant b) \approx F(b) \text{ ili } F(b + \tfrac{1}{2}),$$
$$\mathbf{P}(a \leqslant X \leqslant b) \approx F(b) - F(a) \text{ ili } F(b + \tfrac{1}{2}) - F(a + \tfrac{1}{2}),$$

gdje je crveni izraz *bez korekcije* a plavi izraz *s korekcijom* po neprekidnosti. Obe aproksimacije su korisne, korekcija daje nešto bolju aproksimaciju uzimajući u obzir činjenicu da je binomna razdioba diskretna a normalna razdioba neprekidna.

Korištenjem aproksimacije binomne normalnom razdiobom probajte riješiti:

DZ. Kolika je vjerojatnost da će u **300** bacanja kocke biti postignuto

a) najviše **250** šestica;

b) barem **250** šestica.

Suma i normalna distribucija

Neka su $X_1, \ldots, X_n$ nezavisne jednako distribuirane sluačjne varijable s normalnom distribucijom $N(\mu, \sigma^2)$. Označimo s $S_n = X_1 + \ldots, X_n$.

Suma i normalna distribucija

Neka su $X_1, \ldots, X_n$ nezavisne jednako distribuirane sluačjne varijable s normalnom distribucijom $N(\mu, \sigma^2)$. Označimo s $S_n = X_1 + \ldots, X_n$. Na prošlom predavanju smo komentirali kako je zbroj dvije nezavisne normalne slučajne varijable opet normalna slučajna varijabla. Ista tvrdnja vrijedi i za zbroj konačno mnogo slučajnih varijabli, odnosno možemo zaključiti da je

$$S_n \sim N(?, ?).$$

Suma i normalna distribucija

Neka su $X_1, \ldots, X_n$ nezavisne jednako distribuirane sluačjne varijable s normalnom distribucijom $N(\mu, \sigma^2)$. Označimo s $S_n = X_1 + \ldots, X_n$. Na prošlom predavanju smo komentirali kako je zbroj dvije nezavisne normalne slučajne varijable opet normalna slučajna varijabla. Ista tvrdnja vrijedi i za zbroj konačno mnogo slučajnih varijabli, odnosno možemo zaključiti da je

$$S_n \sim N(?, ?).$$

Odredimo parametre normalne distribucije za S_n. Prvi parametar je

$$\mathbf{E}(S_n) = (\text{linearnost očekivanja}) = \mathbf{E}(X_1) + \ldots + \mathbf{E}(X_n) = \mu + \ldots + \mu = n\mu.$$

Suma i normalna distribucija

Neka su $X_1, \ldots, X_n$ nezavisne jednako distribuirane sluačjne varijable s normalnom distribucijom $N(\mu, \sigma^2)$. Označimo s $S_n = X_1 + \ldots, X_n$. Na prošlom predavanju smo komentirali kako je zbroj dvije nezavisne normalne slučajne varijable opet normalna slučajna varijabla. Ista tvrdnja vrijedi i za zbroj konačno mnogo slučajnih varijabli, odnosno možemo zaključiti da je

$$S_n \sim N(?, ?).$$

Odredimo parametre normalne distribucije za S_n. Prvi parametar je

$$\mathbf{E}(S_n) = (\text{linearnost očekivanja}) = \mathbf{E}(X_1) + \ldots + \mathbf{E}(X_n) = \mu + \ldots + \mu = n\mu.$$

Drugi parametar je

$$\mathrm{Var}(S_n) = (\text{nezavisne sl. varijable}) = \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n) = \sigma^2 + \ldots + \sigma^2 = n\sigma^2.$$

Dakle,

$$S_n \sim N(n\mu, n\sigma^2).$$

Prosjek i normalna distribucija

Neka su kao i prije $X_1, \ldots, X_n$ nezavisne jednako distribuirane sluačjne varijable s normalnom distribucijom $N(\mu, \sigma^2)$. Označimo s $S_n = X_1 + \ldots, X_n$ i $\overline{X}_n = \frac{1}{n} S_n$.

Prosjek i normalna distribucija

Neka su kao i prije $X_1, \ldots, X_n$ nezavisne jednako distribuirane sluačjne varijable s normalnom distribucijom $N(\mu, \sigma^2)$. Označimo s $S_n = X_1 + \ldots, X_n$ i $\overline{X}_n = \frac{1}{n} S_n$. Na prošlom predavanju smo komentirali da ako normalnu slučajnu varijablu pomnožimo s konstantom (ovdje je konstanta $\frac{1}{n}$), dobivena slučajna varijabla i dalje ima normalnu razdiobu. Zaključujemo da je

$$\overline{X}_n \sim N(?, ?).$$

Prosjek i normalna distribucija

Neka su kao i prije $X_1, \ldots, X_n$ nezavisne jednako distribuirane sluačjne varijable s normalnom distribucijom $N(\mu, \sigma^2)$. Označimo s $S_n = X_1 + \ldots, X_n$ i $\overline{X}_n = \frac{1}{n} S_n$. Na prošlom predavanju smo komentirali da ako normalnu slučajnu varijablu pomnožimo s konstantom (ovdje je konstanta $\frac{1}{n}$), dobivena slučajna varijabla i dalje ima normalnu razdiobu. Zaključujemo da je

$$\overline{X}_n \sim N(?, ?).$$

Odredimo parametre normalne distribucije za $\overline{X}_n$. Prvi parametar je

$$\mathbf{E}(\overline{X}_n) = \mathbf{E}(\tfrac{1}{n} S_n) = (\text{linearnost očekivanja}) = \tfrac{1}{n} \mathbf{E}(S_n) = \tfrac{1}{n} \cdot n\mu = \mu.$$

Prosjek i normalna distribucija

Neka su kao i prije $X_1, \ldots, X_n$ nezavisne jednako distribuirane sluačjne varijable s normalnom distribucijom $N(\mu, \sigma^2)$. Označimo s $S_n = X_1 + \ldots, X_n$ i $\overline{X}_n = \frac{1}{n} S_n$. Na prošlom predavanju smo komentirali da ako normalnu slučajnu varijablu pomnožimo s konstantom (ovdje je konstanta $\frac{1}{n}$), dobivena slučajna varijabla i dalje ima normalnu razdiobu. Zaključujemo da je

$$\overline{X}_n \sim N(?, ?).$$

Odredimo parametre normalne distribucije za $\overline{X}_n$. Prvi parametar je

$$\mathbf{E}(\overline{X}_n) = \mathbf{E}(\tfrac{1}{n} S_n) = (\text{linearnost očekivanja}) = \tfrac{1}{n}\mathbf{E}(S_n) = \tfrac{1}{n} \cdot n\mu = \mu.$$

Drugi parametar je

$$\mathrm{Var}(\overline{X}_n) = \mathrm{Var}(\tfrac{1}{n} S_n) = (\text{konst izlazi s kvadratom}) = \tfrac{1}{n^2}\mathrm{Var}(S_n) = \tfrac{1}{n^2} \cdot n\sigma^2 = \tfrac{\sigma^2}{n}.$$

Dakle,

$$\overline{X}_n \sim N(\mu, \tfrac{\sigma^2}{n}).$$

Aproksimacija binomne distribucije normalnom - nastavak

Kakva je veza gornjeg rezultata i de Moivre-Laplaceovog teorema?

Aproksimacija binomne distribucije normalnom - nastavak

Kakva je veza gornjeg rezultata i de Moivre-Laplaceovog teorema? Podsjetimo se da binomnu slučajnu varijablu $X \sim B(n, p)$ možemo prikazati kao sumu n nezavisnih, jednako distribuiranih Bernoullijevih $X_1, \ldots, X_n \sim B(p)$,

$$X = X_1 + \ldots + X_n.$$

Aproksimacija binomne distribucije normalnom - nastavak

Kakva je veza gornjeg rezultata i de Moivre-Laplaceovog teorema? Podsjetimo se da binomnu slučajnu varijablu $X \sim B(n, p)$ možemo prikazati kao sumu n nezavisnih, jednako distribuiranih Bernoullijevih $X_1, \ldots, X_n \sim B(p)$,

$$X = X_1 + \ldots + X_n.$$

Također, sjetimo se da je za Bernoullijevu radiobu očekivanje jednako $\mu = p$ i varijanca jednaka $\sigma^2 = p(1-p)$.

Aproksimacija binomne distribucije normalnom - nastavak

Kakva je veza gornjeg rezultata i de Moivre-Laplaceovog teorema? Podsjetimo se da binomnu slučajnu varijablu $X \sim B(n, p)$ možemo prikazati kao sumu n nezavisnih, jednako distribuiranih Bernoullijevih $X_1, \ldots, X_n \sim B(p)$,

$$X = X_1 + \ldots + X_n.$$

Također, sjetimo se da je za Bernoullijevu radiobu očekivanje jednako $\mu = p$ i varijanca jednaka $\sigma^2 = p(1-p)$.
Znamo da za velike n možemo X aproksimirati odgovarajućom normalnom razdiobom

$$X \approx N(n\mu, n\sigma^2).$$

Uočite da je to ista razdioba kao i za sumu S_n n nezavisnih jednako-distribuiranih normalnih slučajnih varijabli s očekivanjem μ i varijancom σ^2.

Aproksimacija binomne distribucije normalnom - nastavak

Kakva je veza gornjeg rezultata i de Moivre-Laplaceovog teorema? Podsjetimo se da binomnu slučajnu varijablu $X \sim B(n, p)$ možemo prikazati kao sumu n nezavisnih, jednako distribuiranih Bernoullijevih $X_1, \ldots, X_n \sim B(p)$,

$$X = X_1 + \ldots + X_n.$$

Također, sjetimo se da je za Bernoullijevu radiobu očekivanje jednako $\mu = p$ i varijanca jednaka $\sigma^2 = p(1-p)$.
Znamo da za velike n možemo X aproksimirati odgovarajućom normalnom razdiobom

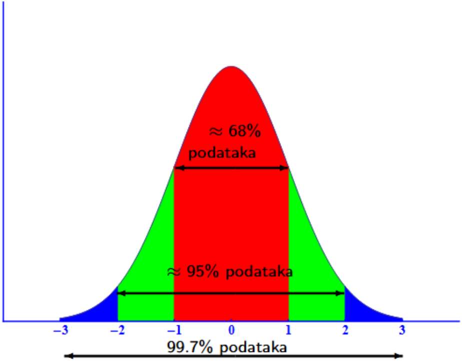
$$X \approx N(n\mu, n\sigma^2).$$

Uočite da je to ista razdioba kao i za sumu S_n n nezavisnih jednako-distribuiranih normalnih slučajnih varijabli s očekivanjem μ i varijancom σ^2. Analogno, kada bi nas zanimao **prosječan broj uspjeha** u n nez. ponavljanja pokusa, $\overline{X_n} = \frac{1}{n}X$, slijedilo bi istim računom kao i ranije da je

$$\overline{X}_n \approx N(\mu, \frac{\sigma^2}{n}).$$

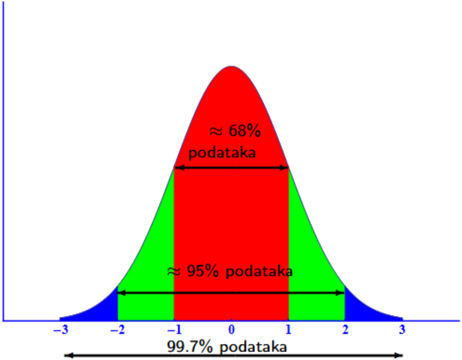
3σ - pravilo za normalnu distribuciju

Neka je Z standardna normalna slučajna varijabla, promotrimo njenu funkciju gustoće:



3σ - pravilo za normalnu distribuciju

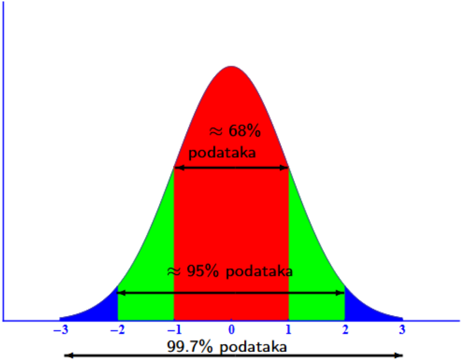
Neka je Z standardna normalna slučajna varijabla, promotrimo njenu funkciju gustoće:



Uočimo da se približno 68% realizacija slučajne varijable Z nalazi u intervalu $[-1, 1]$ (crvena površina):

3σ - pravilo za normalnu distribuciju

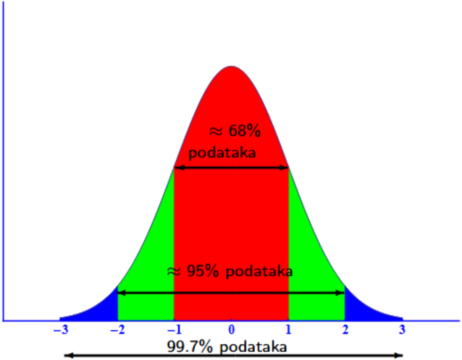
Neka je Z standardna normalna slučajna varijabla, promotrimo njenu funkciju gustoće:



Uočimo da se približno 68% realizacija slučajne varijable Z nalazi u intervalu $[-1, 1]$ (crvena površina): $\mathbf{P}(|Z| \leqslant 1) = \mathbf{P}(-1 \leqslant Z \leqslant 1) \approx \mathit{red}0.68$.

3σ - pravilo za normalnu distribuciju

Neka je Z standardna normalna slučajna varijabla, promotrimo njenu funkciju gustoće:



Uočimo da se približno 68% realizacija slučajne varijable Z nalazi u intervalu $[-1, 1]$ (crvena površina): $\mathbf{P}(|Z| \leqslant 1) = \mathbf{P}(-1 \leqslant Z \leqslant 1) \approx red0.68$.

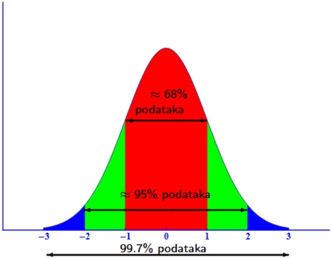
Analogno: $\mathbf{P}(-2 \leqslant Z \leqslant 2) \approx 0.95$, $\mathbf{P}(-3 \leqslant Z \leqslant 3) \approx 0.997$.

3σ - pravilo za normalnu distribuciju

Neka je $X \sim N(\mu, \sigma^2)$ normalna slučajna varijabla. Tada je graf funkcija gustoće slučajne varijable

$$\frac{X-\mu}{\sigma} \sim N(0,1)$$

dan s:

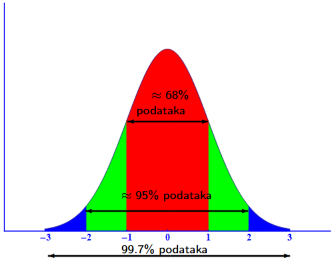


3σ - pravilo za normalnu distribuciju

Neka je $X \sim N(\mu, \sigma^2)$ normalna slučajna varijabla. Tada je graf funkcija gustoće slučajne varijable

$$\frac{X - \mu}{\sigma} \sim N(0, 1)$$

dan s:



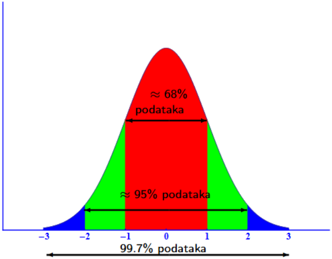
Uočimo da se približno 68% realizacija slučajne varijable X nalazi u intervalu $[\mu - \sigma, \mu + \sigma]$ (crvena površina):

3σ - pravilo za normalnu distribuciju

Neka je $X \sim N(\mu, \sigma^2)$ normalna slučajna varijabla. Tada je graf funkcija gustoće slučajne varijable

$$\frac{X-\mu}{\sigma} \sim N(0,1)$$

dan s:



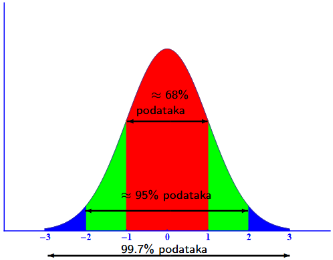
Uočimo da se približno 68% realizacija slučajne varijable X nalazi u intervalu $[\mu-\sigma, \mu+\sigma]$ (crvena površina): $\mathbf{P}\left(\left|\frac{X-\mu}{\sigma}\right| \leqslant 1\right) = \mathbf{P}(\mu-\sigma \leqslant X \leqslant \mu+\sigma) \approx 0.68$.

3σ - pravilo za normalnu distribuciju

Neka je $X \sim N(\mu, \sigma^2)$ normalna slučajna varijabla. Tada je graf funkcija gustoće slučajne varijable

$$\frac{X-\mu}{\sigma} \sim N(0,1)$$

dan s:



Uočimo da se približno 68% realizacija slučajne varijable X nalazi u intervalu $[\mu-\sigma, \mu+\sigma]$ (crvena površina): $\mathbf{P}\left(\left|\frac{X-\mu}{\sigma}\right| \leqslant 1\right) = \mathbf{P}(\mu-\sigma \leqslant X \leqslant \mu+\sigma) \approx 0.68$.

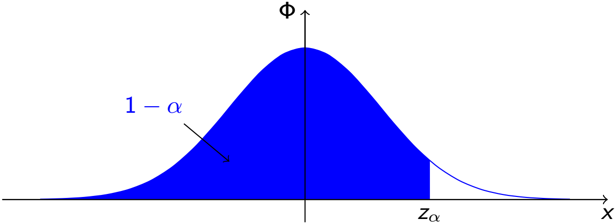
Analogno: $\mathbf{P}(\mu-2\sigma \leqslant X \leqslant \mu+2\sigma) \approx 0.95$, $\mathbf{P}(\mu-3\sigma \leqslant Z \leqslant \mu+3\sigma) \approx 0.997$.

Kvantili normalne distribucije

Općenito, za $\alpha \in (0, 1)$, broj z_α za koji vrijedi da je

$$F_{N(0,1)}(z_\alpha) = \Phi(z_\alpha) = 1 - \alpha$$

zovemo $(1 - \alpha)$-kvantil standardne normalne razdiobe.

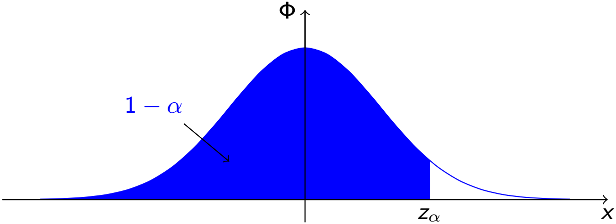


Kvantili normalne distribucije

Općenito, za $\alpha \in (0, 1)$, broj z_α za koji vrijedi da je

$$F_{N(0,1)}(z_\alpha) = \Phi(z_\alpha) = 1 - \alpha$$

zovemo $(1 - \alpha)$-kvantil standardne normalne razdiobe.



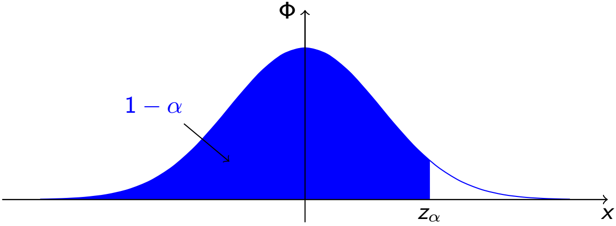
Kako računamo z_α?

Kvantili normalne distribucije

Općenito, za $\alpha \in (0,1)$, broj z_α za koji vrijedi da je

$$F_{N(0,1)}(z_\alpha) = \Phi(z_\alpha) = 1 - \alpha$$

zovemo $(1-\alpha)$-kvantil standardne normalne razdiobe.



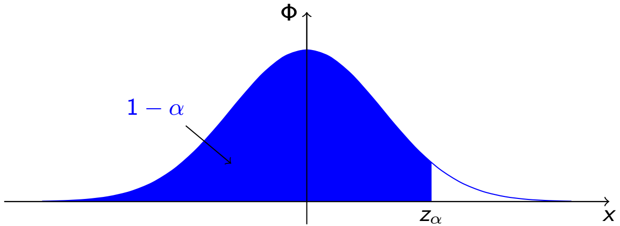
Kako računamo z_α? Uočimo da je $z_\alpha = \Phi^{-1}(1-\alpha)$, gdje je Φ^{-1} inverz funkcije Φ.

Kvantili normalne distribucije

Općenito, za $\alpha \in (0,1)$, broj z_α za koji vrijedi da je

$$F_{N(0,1)}(z_\alpha) = \Phi(z_\alpha) = 1 - \alpha$$

zovemo $(1-\alpha)$-kvantil standardne normalne razdiobe.



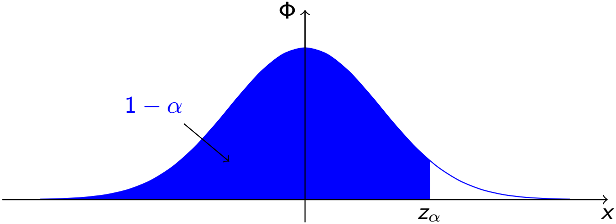
Kako računamo z_α? Uočimo da je $z_\alpha = \Phi^{-1}(1-\alpha)$, gdje je Φ^{-1} inverz funkcije Φ. U Excelu koristimo funkciju `= NORM.INV(1-`α`, 0, 1)`.

Kvantili normalne distribucije

Općenito, za $\alpha \in (0,1)$, broj z_α za koji vrijedi da je

$$F_{N(0,1)}(z_\alpha) = \Phi(z_\alpha) = 1 - \alpha$$

zovemo $(1-\alpha)$-kvantil standardne normalne razdiobe.



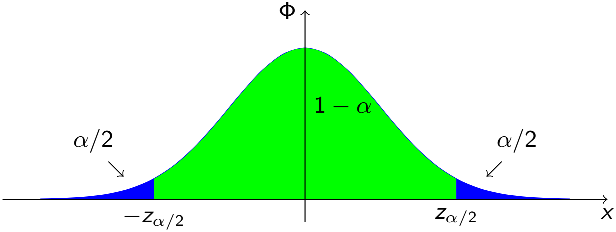
Kako računamo z_α? Uočimo da je $z_\alpha = \Phi^{-1}(1-\alpha)$, gdje je Φ^{-1} inverz funkcije Φ. U Excelu koristimo funkciju `= NORM.INV(1-`$\alpha, 0, 1$`)`.

Na primjer $z_{0.05}$ = `NORM.INV(0.95, 0, 1)`$= 1.645$ i
$z_{0.025}$ = `NORM.INV(0.975, 0, 1)`$= 1.96$.

Kvantili standardne normalne razdiobe - simetrija

Neka je $Z \sim N(0,1)$. Uočimo da je zbog simetrije standardne normalne razdiobe

$$\mathbf{P}(-z_{\alpha/2} \leqslant Z \leqslant z_{\alpha/2}) = 1 - \alpha,$$



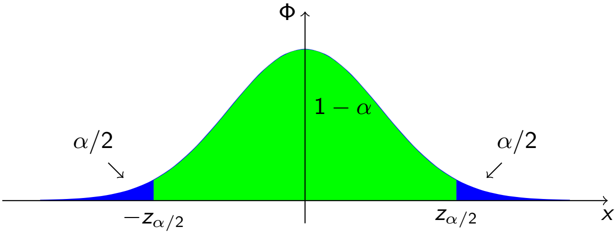
odnosno za $X \sim N(\mu, \sigma^2)$

$$\mathbf{P}(\mu - z_{\alpha/2}\sigma \leqslant X \leqslant \mu + z_{\alpha/2}\sigma) = 1 - \alpha.$$

Kvantili standardne normalne razdiobe - simetrija

Neka je $Z \sim N(0,1)$. Uočimo da je zbog simetrije standardne normalne razdiobe

$$\mathbf{P}(-z_{\alpha/2} \leqslant Z \leqslant z_{\alpha/2}) = 1 - \alpha,$$



odnosno za $X \sim N(\mu, \sigma^2)$

$$\mathbf{P}(\mu - z_{\alpha/2}\sigma \leqslant X \leqslant \mu + z_{\alpha/2}\sigma) = 1 - \alpha.$$

Na primjer $z_{0.025} = 1.96$ pa gornji izraz za $\alpha = 0.05$ postaje

$$\mathbf{P}(\mu - 1.96 \cdot \sigma \leqslant X \leqslant \mu + 1.96 \cdot \sigma) = 0.95.$$

Kako je $1.96 \approx 2$ vidimo da je dana jednakost osnova za 2σ-pravilo.

Još neke neprekidne distribucije

- Studentova (t) distribucija
- Fischerova (F) distribucija
- 'chi' kvadrat (χ^2) distribucija