7 Vztahy mezi jednotlivými rozděleními

Teorie: Vztahy mezi rozděleními

http://www.johndcook.com/blog/distribution_chart/
http://www.math.wm.edu/~leemis/2008amstat.pdf

$A(p)$	$=$	$Bi(n=1,p)$	Bernoulliho rozdělení je speciální případ binomického
$HG(N,K,n)$	$\approx$	$Bi\left(p=\dfrac{K}{N};n\right)$	pro $n\to\infty$ a $\dfrac{K}{N}\to 0$
$Bi(n,p)$	$\approx$	$Po(\lambda=n\,p)$	pro $n\to\infty$ a $p\to 0$ (stačí $n\geq 30$ a $p\leq 0.1$)
$Bi(n,p)$	$\approx$	$N(np,np(1-p))$	pro $np(1-p)\geq 9$
$Po(\lambda)$	$\approx$	$N(\lambda,\lambda)$	pro $\lambda\geq 9$

(7.1) Ze zásilky velikosti 100 výrobků, která obsahuje 5 vadných vybíráme 10 kusů. Určete pravděpodobnost, že ve výběru bude nejvýše jeden vadný

(a) pomocí hypergeometrického rozdělení

[$X\sim HG(K=5,N=100,n=10)$]
[$P(X\leq 1)=P(0)+P(1)=0.9231$]

(b) pomocí binomického rozdělení

[$X\sim HG(K=5,N=100,n=10)\approx Bi(p=5/100,n=10)$]
[$P(X\leq 1)=P(0)+P(1)=0.9139$]

(c) pomocí Poissonova rozdělení

[$X\sim HG(K=5,N=100,n=10)\approx Bi(p=5/100,n=10)\approx Po(\lambda=5/100\cdot 10$]
[$P(X\leq 1)=P(0)+P(1)=F(1)=0.9098$]

(7.2) Pravděpodobnost překlepu na stránce je 10%. Určete pravděpodobnost, že ze 100 stránek textu budou nejvýše 4 strany s překlepem.

(a) pomocí binomického rozdělení

[$X\sim Bi(p=10\%,n=100)$]
[$P(X\leq 4)=P(0)+P(1)+P(2)+P(3)+P(4)=0.0237$]

(b) aproximace binomického rozdělení normálním

[$X\sim Bi(p=10\%,n=100)\approx N(\mu=10,\sigma^2=9)$]
[$P(X\leq 4)=F(4)=0.0228$]

(7.3) Pro $X\sim Po(\lambda=10)$ určete $P(X\leq 13)$

(a) pomocí tabulek Poissonova rozdělení

[$P(X\leq 13)=F(13)=0.8645$]

(b) aproximace Poissonova rozdělení normálním

[$X\sim Po(\lambda=10)\approx N(\mu=10,\sigma^2=10)$]
[$P(X\leq 13)=F(13)=\Phi\left(\dfrac{13-10}{\sqrt{10}}\right)=0.8286$]

Teorie: Rozdělení součtu nezávislých náhodných veličin

$X_i, i = 1, 2, \dots, n \sim$ Bernoulliho $A(p)$	$\dots$	$\sum_{i=1}^{n} X_i \sim$ binomické $Bi(n, p)$
$X_1 \sim Po(\lambda_1)$, $X_2 \sim Po(\lambda_2)$	$\dots$	$X_1 + X_2 \sim Po(\lambda_1 + \lambda_2)$
$X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$	$\dots$	$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$
$X_1, X_2 \sim$ rovnoměrné	$\dots$	$X_1 + X_2 \sim$ trojúhelníkové
$X_1 \sim f_1$, $X_2 \sim f_2$	$\dots$	$X_1 + X_2 \sim f(y) = \int f_1(x) f_2(y - x) dx$

Použití normálního rozdělení pro součty a průměry (centrální limitní věta)

$$\sum_{i=1}^{n} X_i \approx N(n\mu, n\sigma^2) \quad \text{pro } X_i \text{ takové, že } \mathrm{E}X_i = \mu, \mathrm{var}X_i = \sigma^2 \text{ (centralní limitní věta)}$$

$$1/n \sum_{i=1}^{n} X_i \approx N(\mu, \sigma^2/n) \quad \text{pro } X_i \text{ takové, že } \mathrm{E}X_i = \mu, \mathrm{var}X_i = \sigma^2 \text{ (CLV pro průměr)}$$

(7.4) Necht' $X \sim R(-1, 1)$ určete rozdělení veličin $Y = 2 \cdot X$ a $Y = X + X$.

(7.5) Telefonní ústředna spojí průměrně 76 hovorů za minutu a jejich počet se řídí Poissonovým rozdělením. Spočtěte pravděpodobnost, že ústředna za minutu spojí více než 80 hovorů.

$$[P(X > 80) = 1 - F_{Poisson}(80) \doteq 1 - F_{Norm.}(80) = 1 - \phi\left(\frac{80 - 76}{\sqrt{76}}\right) = 1 - 0.6768 = 0.3232\ ]$$

(7.6) Zaokrouhlovací chyba na celé jednotky má rovnoměrné rozložení na intervalu (-0.5; 0.5). Spočtěte pravděpodobnost, že součet 100 zaokrouhlovacích chyb (nezávislých) bude v absolutní hodnotě menší než 5.

$$[\ X_i \sim R(-0.5; 0.5) \Rightarrow S = \sum_{i=1}^{100} X_i \sim N\left(100 \cdot 0; \frac{100}{12}\right)]$$

$$[\ P(-5 < S < 5) = F(5) - F(-5) = \phi\left(\frac{5}{\sqrt{\frac{100}{12}}}\right) - 1 + \phi\left(\frac{5}{\sqrt{\frac{100}{12}}}\right) = 0.9164\ ]$$

(7.7) Určete pravděpodobnost, že při 100-krát opakovaném hodu kostkou padne šestka nejvýše 20-krát.

[počet šestek $X \sim Bi(n = 100, p = 1/6) \Rightarrow X \sim N(\mu = 100 \cdot 1/6, \sigma^2 = 100 \cdot 1/6 \cdot 5/6)$]

[$P(X \leq 20) = F(20) = \phi(0.8944) = 81.44\%$]

(7.8) Měsíční náklady na opravu zařízení mají normální rozdělení se střední hodnotou 1000 Kč a směrodatnou odchylkou 120 Kč.

(a) Určete pravděpodobnost, že měsíční náklady jsou větší než 1100 Kč.
$[\ Naklady \sim N(\mu = 1000, \sigma^2 = 120^2) \Rightarrow P(X > 1100) = 1 - F(1100) = \phi(5/6) = 0.2023\ ]$

(b) Určete pravděpodobnost, že průměrné měsíční náklady za poslední rok (12 měsíců) jsou větší než 1100 Kč.

$$[\ X = \frac{1}{12}\sum_{i=1}^{12} N_i \Rightarrow S \sim N(\mu = 1000, \sigma^2 = 120^2/12)]$$

$$[P(X > 1100) = 1 - F(1100) = \phi(2.89) = 0.0019\ ]$$

(7.9) Součástky s délkou $X \sim N(\mu = 150\,\text{mm}; \sigma^2 = 2.53\,\text{mm}^2)$ se balí do krabic, které mají délkou $Y \sim N(\mu = 155\ \text{mm}; \sigma^2 = 0.36\,\text{mm}^2)$. Balíme 3000 součástek za směnu. Linka se zastaví pokud je součástka delší než krabice.

(a) Určete pravděpodobnost, součástka se nevejde do krabice (bude delší).
$[\ \text{Rozdíl}\sim N(\mu = 155 - 150, \sigma^2 = 2.53 + 0.36 = 2.89), P(\text{Rozdíl} < 0) = F(0) = 0.16\%\ ]$

(b) Určete ppst., že se linka zastaví během směny alespoň jednou.
$[\text{Počet zastavení } X \sim Bi(n = 3000, p = 0.16\%), P(X \geq 1) = 1 - P(0) = 99.18\%\ ]$

(c) Určete průměrný počet zastavení linky během směny.
$[\ EX = n \cdot p = 3000 \cdot 0.0016 = 4.8\ ]$

(7.10) Sčítáme 100 reálných čísel, které jsme před sčítáním zaokrouhlili na čísla celá. Určete s 90% ppstí chybu $\pm\varepsilon$, které jsme se dopustili zaokrouhlováním.

$$[\text{chyba} \sim R(-1/2, 1/2), \text{součet chyb}\, X \sim N(0, 100 \cdot 1/12)\ ]$$

$$[P(-\varepsilon < \text{součet chyb} < \varepsilon) = 0.9 \Rightarrow \varepsilon = 4.75\ ]$$

(7.11) Povolená hmotnostní odchylka pro spotřebitelská balení suchých skořápkových plodů do 100 g je ± 5 g. Bylo zjištěno, že hmotnost krabic, které obsahují 20 balení se řídí normálním rozdělením s parametry $\mu = 2000$ g a $\sigma^2 = 200$ g^2. Určete směrodatnou odchylku jednoho 100 gramového balení a pravděpodobnost, že toto balení splňuje povolené hmotnostní odchylky.

$$[\sum_{i=1}^{20} X_i \sim N(\mu = 2000, \sigma^2 = 200) \Rightarrow X_i \sim N(\mu = 100, \sigma^2 = 10)\ ]$$

$$[P(95 < X_i < 105) = F(105) - F(95) = \phi(1.58) - \phi(-1.58) = 88.62\%\ ]$$

(7.12) Dlouhodobým pozorováním bylo zjištěno, že doba potřebná k objevení a odstranění poruchy je náhodná veličina se střední hodnotou 40 minut a směrodatnou odchylkou 30 minut. Určete jakou dobu si vyžádá objevení a odstranění 100 poruch, jestliže žádáme, aby tato hranice nebyla s pravděpodobností 0.95 překročena.

$$[\ \sum_{i=1}^{100} X \sim N(\mu = 100 \cdot 40; \sigma^2 = 100 \cdot 30^2)\ ]$$

$$[\ \text{hledám 95\% kvantil } x_{0.95} = \mu + \sigma \cdot u_{0.95} = 4000 + \sqrt{90000} \cdot 1.645 = 4493.5 \text{ minut}\ ]$$