

Hoofdstuk 6

Discrete distributies

Marnix Van Daele

Marnix.VanDaele@UGent.be

Vakgroep Toegepaste Wiskunde en Informatica
Universiteit Gent

Discrete distributies

- binomiale verdeling
- Poisson verdeling
- hypergeometrische verdeling
- uniforme discrete verdeling
- ...

De binomiale verdeling

De binomiaal verdeelde toevalsveranderlijke I is het aantal keer dat een verschijnsel A optreedt in een reeks van n enkelvoudige waarnemingen. Hierbij moet de kans op het optreden van een verschijnsel A bij een enkelvoudige waarneming de constante waarde θ bedragen :

$$P(A) = \theta .$$

Voorbeeld : zij I is het aantal keer dat 5 gegooit wordt bij 7 onafhankelijke worpen met een dobbelsteen.

$$I \text{ is binomiaal verdeeld met } \begin{cases} n = 7 \\ \theta_I = \frac{1}{6} \end{cases}$$

Distributie

$$\begin{aligned}
 & P(\underbrace{A \quad A \quad \dots \quad A}_i \underbrace{\overline{A} \quad \overline{A} \quad \dots \quad \overline{A}}_{n-i}) \\
 &= \underbrace{\theta \quad \theta \quad \dots \quad \theta}_i \underbrace{(1-\theta) (1-\theta) \dots (1-\theta)}_{n-i} = \theta^i (1-\theta)^{n-i}
 \end{aligned}$$

$$\begin{aligned}
 & P(\underbrace{\overline{A} \quad \overline{A} \quad \dots \quad \overline{A}}_{n-i} \underbrace{A \quad A \quad \dots \quad A}_i) \\
 &= \underbrace{(1-\theta) (1-\theta) \dots (1-\theta)}_{n-i} \underbrace{\theta \quad \theta \quad \dots \quad \theta}_i = \theta^i (1-\theta)^{n-i}
 \end{aligned}$$

aantal mogelijke sequenties : $C_n^i = \frac{n!}{i! (n-i)!} = \binom{n}{i}$

elke sequentie heeft kans $\theta^i (1-\theta)^{n-i}$

Besluit : $\varphi_I(i) = \binom{n}{i} \theta^i (1-\theta)^{n-i} \quad i = 0, 1, \dots, n$

Distributie

$$\varphi_I(i) = \binom{n}{i} \theta^i (1 - \theta)^{n-i} \quad i = 0, 1, \dots, n$$

$$\Phi_I(w) = \mathbf{P}(I \leq w) = \sum_{i \leq w} \binom{n}{i} \theta^i (1 - \theta)^{n-i}$$

$$\text{binomium van Newton : } (a + b)^n = \sum_{i=0}^n \binom{n}{i} a^i b^{n-i}$$

$$\Phi_I(n) = \sum_{i=0}^n \binom{n}{i} \theta^i (1 - \theta)^{n-i} = [\theta + (1 - \theta)]^n = 1$$

Karakteristieken

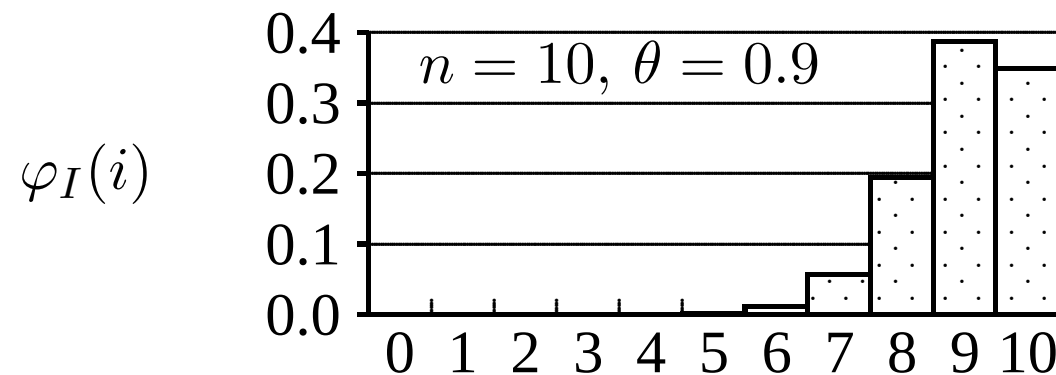
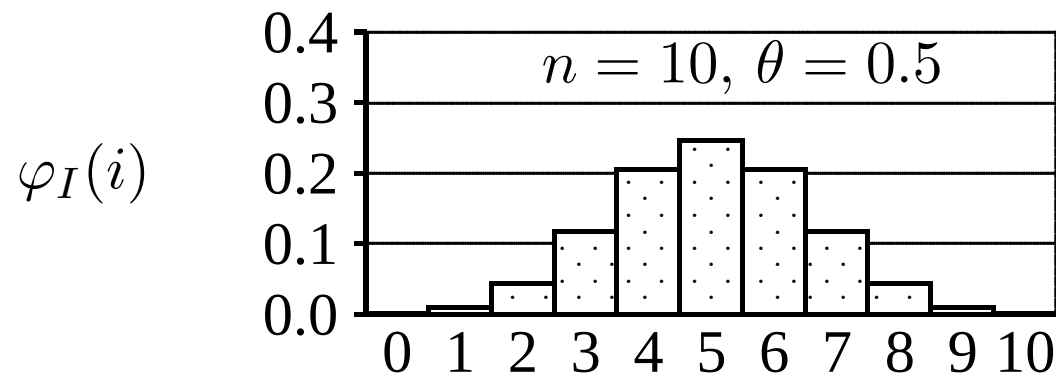
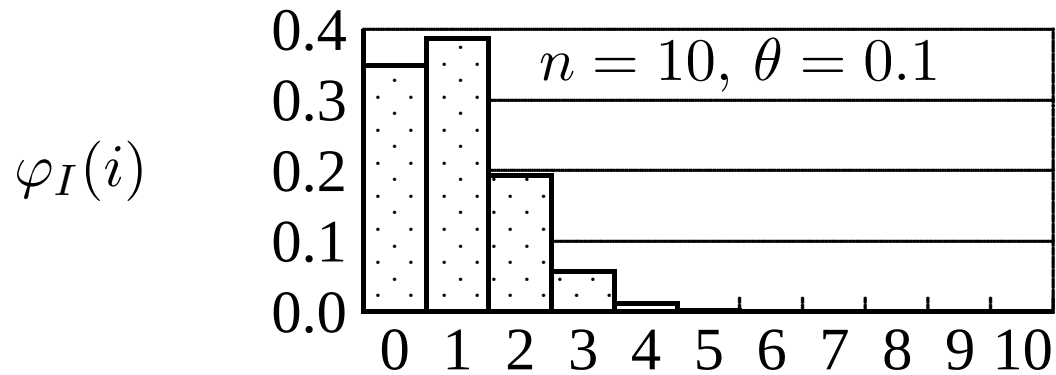
$$\mu_I = \sum_{i=0}^n i \varphi_I(i) = n \theta$$

$$\sigma_I^2 = \sum_{i=0}^n i^2 \varphi_I(i) - \mu^2 = n \theta (1 - \theta)$$

$$\sigma_I = \sqrt{n \theta (1 - \theta)}$$

θ : parameter van de binomiale distributie

Kansverdeling



Voorbeeld

Een familie heeft 6 zes kinderen. De kansen bij een geboorte bedroegen 0.49 voor een jongen en 0.51 voor een meisje.

- (i) Wat is de kans dat er tenminste 1 meisje is ?
- (ii) Wat is de kans dat er hoogstens 2 jongens zijn ?

Oplossing : Zij I het aantal meisjes onder de zes kinderen.

$$\varphi_I(i) = \binom{6}{i} \theta^i (1 - \theta)^{6-i} \quad \theta = 0.51$$

$$\begin{aligned} \text{(i)} \quad \mathbf{P}(I \geq 1) &= 1 - \mathbf{P}(I < 1) = 1 - \mathbf{P}(I = 0) = 1 - \varphi_I(0) \\ &= 1 - (1 - \theta)^6 = 0.9826 \end{aligned}$$

$$\begin{aligned} \text{(ii)} \quad \mathbf{P}(I \geq 4) &= \varphi_I(4) + \varphi_I(5) + \varphi_I(6) \\ &= \binom{6}{4} \theta^4 (1 - \theta)^2 + \binom{6}{5} \theta^5 (1 - \theta) + \theta^6 = 0.3627 \end{aligned}$$

Toepassing

Zij $x_1, x_2, \dots, x_{10}$ 10 onafhankelijke waarden van de toevalsveranderlijke X met cumulatieve distributiefunctie $\Phi_X(x)$. Bepaal de kans dat alle meetwaarden vallen in $[\alpha, \beta]$.

Oplossing : $P(\alpha \leq X \leq \beta) = \Phi_X(\beta) - \Phi_X(\alpha)$

Zij I het aantal x -en in $[\alpha, \beta]$.

Dan is I binomiaal verdeeld met
$$\begin{cases} n = 10 \\ \theta_I = \Phi_X(\beta) - \Phi_X(\alpha) \end{cases}$$

$$P(I = 10) = \theta_I^{10} = (\Phi_X(\beta) - \Phi_X(\alpha))^{10}$$

Teruglegging

Voorbeeld : Een urne bevat $n_U = 100$ (op kleur na identieke) ballen, waaronder $k = 20$ zwarte. We voeren $n = 25$ achtereenvolgende trekkingen van een bal uit. Zij I het aantal getrokken zwarte ballen.

- met teruglegging : de kans op een zwarte bal is voor elke trekking constant, nl. $\frac{k}{n_U}$

I : binomiaal verdeeld met $\theta_I = \frac{k}{n_U}$

- zonder teruglegging : de kans op een zwarte bal verschilt van trekking tot trekking

I : hypergeometrisch verdeeld

Hypergeometrische distributie

De hypergeometrisch verdeelde toevalsveranderlijke I is het aantal keer dat een eigenschap A wordt waargenomen in een reeks van n enkelvoudige waarnemingen van telkens verschillende elementen uit een verzameling van n_U elementen waarvan k elementen deze eigenschap A bezitten.

Distributie

- n_U : populatiegrootte
- n : steekproefgrootte
- k : aantal elementen in populatiemet gezochte eigenschap

$$\max(0, n + k - n_U) \leq I \leq \min(n, k)$$

want $I \leq k$ en als $n \geq n_U - k$, dan is $I \geq n - (n_U - k)$

$$\varphi_I(i) = \frac{C_k^i C_{n_U-k}^{n-i}}{C_{n_U}^n} = \frac{\binom{k}{i} \binom{n_U-k}{n-i}}{\binom{n_U}{n}}$$

Karakteristieken

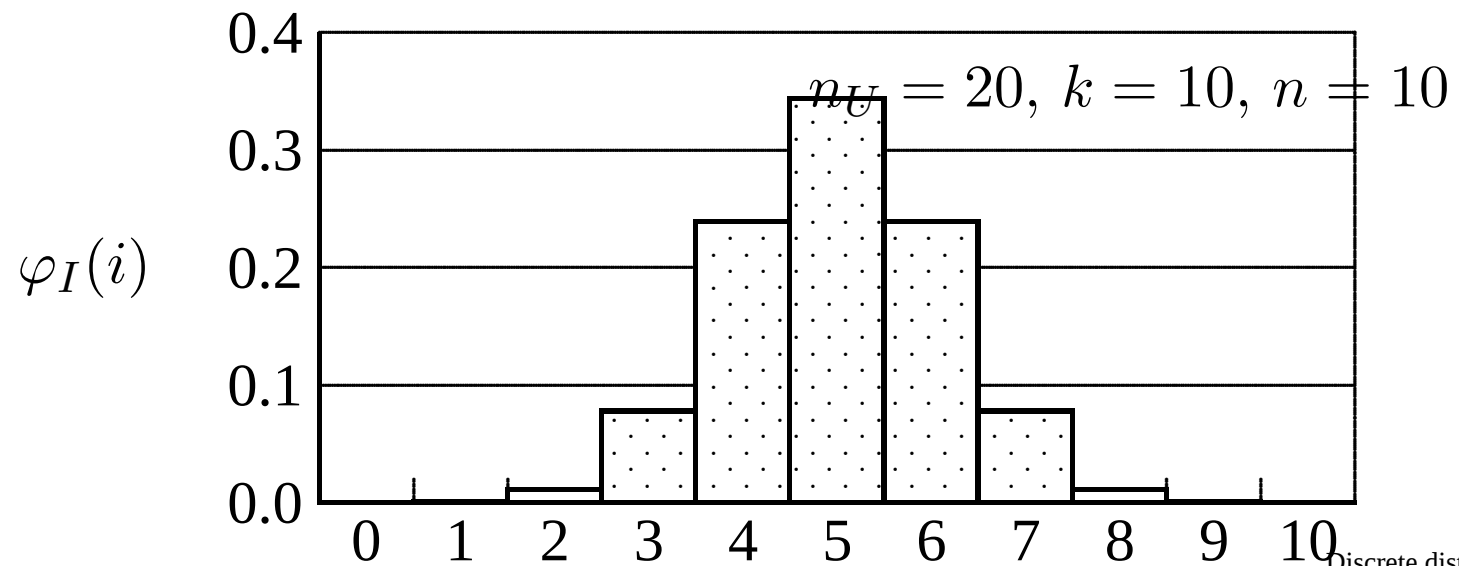
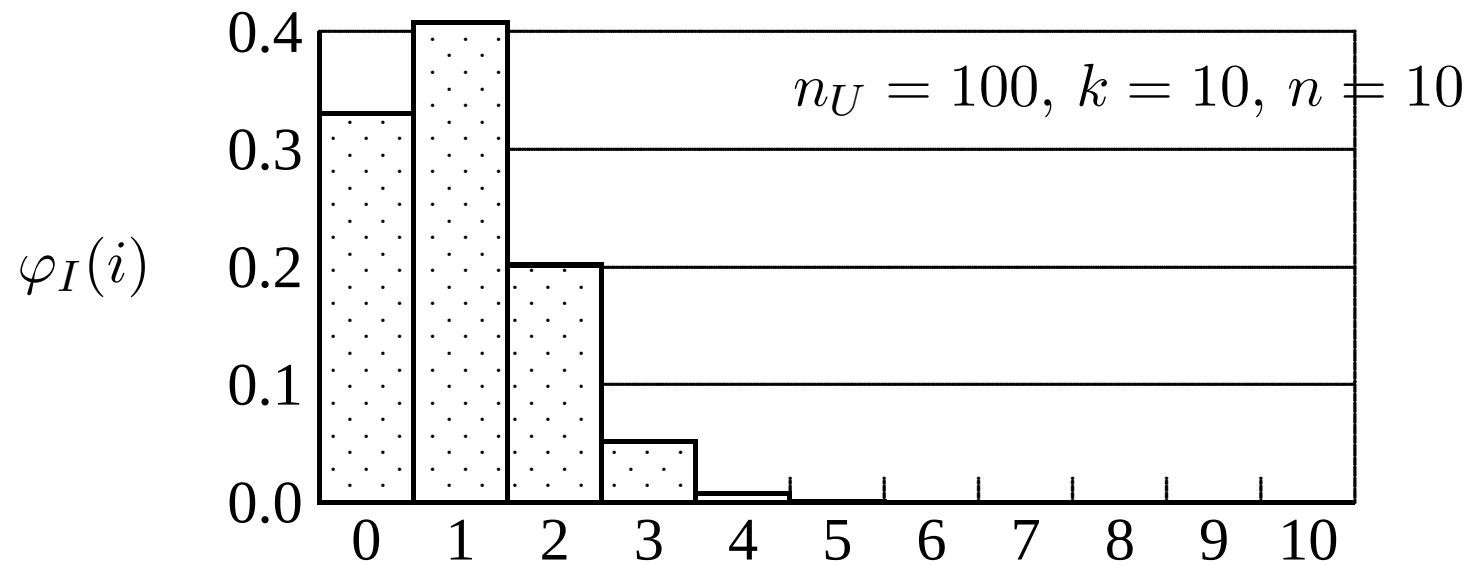
- n_U : populatiegrootte
- n : steekproefgrootte
- k : aantal elementen in populatiemet gezochte eigenschap

$$\mu_I = n \frac{k}{n_U} = n \theta \quad \theta = \frac{k}{n_U}$$

$$\sigma_I^2 = n \frac{k}{n_U} \frac{n_U - k}{n_U} \frac{n_U - n}{n_U - 1} = n \theta (1 - \theta) \left(1 - \frac{n - 1}{n_U - 1} \right)$$

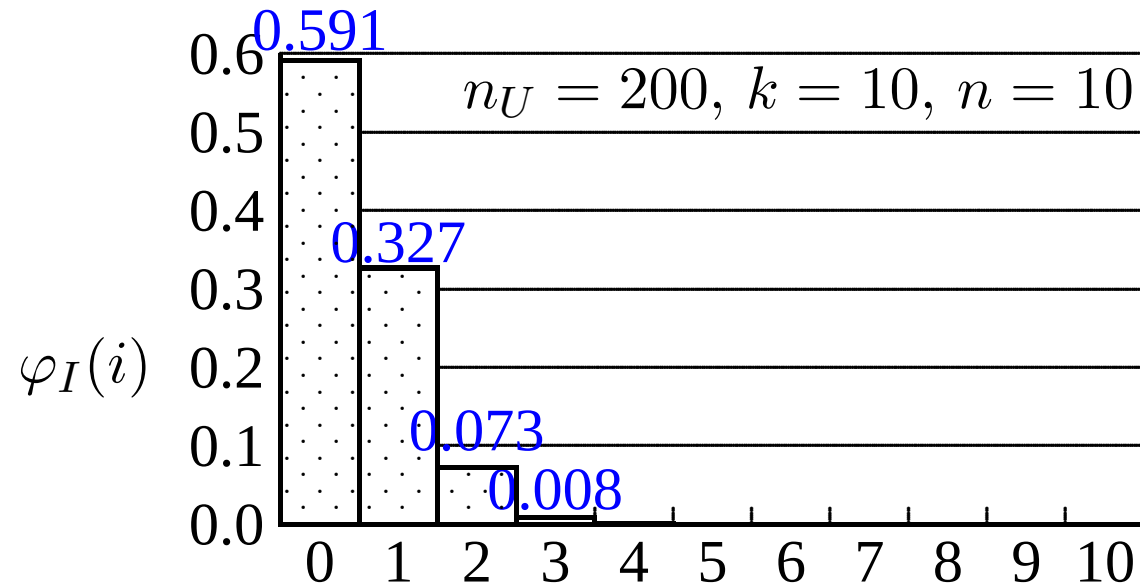
Als $\frac{n - 1}{n_U - 1} \approx \frac{n}{n_U}$ zeer klein is (d.z.w. ongeveer 0), kan de hypergeometrische verdeling benaderd worden door een binomiale verdeling met parameter $\theta = \frac{k}{n_U}$

Kansverdeling

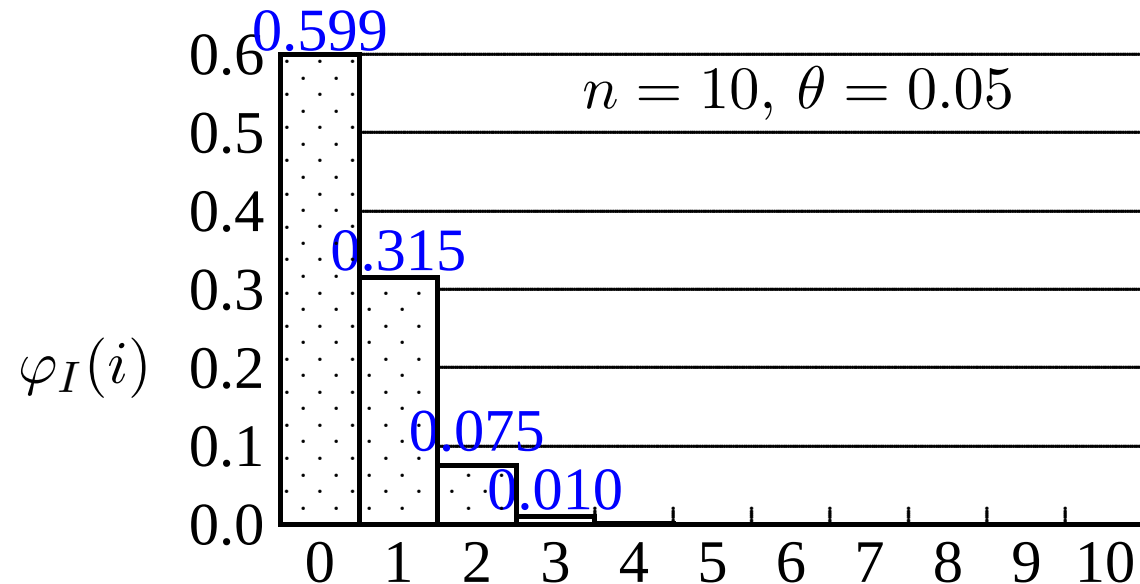


Vergelijking

Hyper-
geometrisch



Binomiaal



Voorbeeld

In een loterij worden, benevens heel wat troostprijzen, 10 hoofdprijzen uitgedeeld. Als er 5000 deelnemende nummers zijn, wat is dan de kans dat een deelnemer met 10 loten minstens 1 hoofdprijs heeft gewonnen.

Oplossing : I : aantal gewonnen hoofdprijzen met de 10 loten.

I : hypergeom. verdeeld met $k = 10$, $n = 10$ en $n_U = 5000$

$$\begin{aligned} P(I \geq 1) &= 1 - P(I = 0) \\ &= 1 - \frac{\binom{10}{0} \binom{4990}{10}}{\binom{5000}{10}} = \frac{(4990!)^2}{4980! 5000!} = 0.019839 \end{aligned}$$

$$\frac{n}{n_U} = 0.002 \implies \text{binomiale benadering met } \theta = \frac{k}{n_U} = 0.002$$

$$P(I \geq 1) = 1 - P(I = 0) \approx 1 - (1 - \theta)^{10} = 0.019821$$

De verdeling van Poisson

De **Poisson verdeelde veranderlijke** I is het aantal keer dat een verschijnsel A optreedt in een totale tijdsduur t . Hierbij moet de kans dat het verschijnsel optreedt in een klein tijdsinterval Δt evenredig zijn met de duur van dit interval : $\lambda \Delta t$. Daarenboven moet het ene optreden van A onafhankelijk zijn van vorige optredens van A , hetgeen betekent dat λ een constante is die niet afhangt van wat voordien voorgevallen is.

klein betekent : het verschijnsel kan hoogstens 1 keer optreden

Distributie

$P_i(t) = \text{P}(\text{A treedt } i \text{ keer op in tijdsduur } t)$

$$P_0(\Delta t) = 1 - P_1(\Delta t) = 1 - \lambda \Delta t$$

- $i = 0$

$$P_0(t + \Delta t) = P_0(t) P_0(\Delta t) = P_0(t) (1 - \lambda \Delta t)$$

$$\Longleftrightarrow P_0(t + \Delta t) - P_0(t) + \lambda P_0(t) \Delta t = 0$$

$$\Longleftrightarrow \lim_{\Delta t \rightarrow 0} \frac{P_0(t + \Delta t) - P_0(t)}{\Delta t} + \lambda P_0(t) = 0$$

$$\Rightarrow \frac{dP_0(t)}{dt} + \lambda P_0(t) = 0 \quad P_0(0) = 1$$

$$P_0(t) = e^{-\lambda t}$$

Distributie

$P_i(t) = \text{P}(\text{A treedt } i \text{ keer op in tijdsduur } t)$

$$P_0(\Delta t) = 1 - P_1(\Delta t) = 1 - \lambda \Delta t$$

- $i > 0$

$$\begin{aligned} P_i(t + \Delta t) &= P_i(t) P_0(\Delta t) + P_{i-1}(t) P_1(\Delta t) \quad i = 1, 2, \dots \\ &= P_i(t) (1 - \lambda \Delta t) + P_{i-1}(t) \lambda \Delta t \end{aligned}$$

$$\Leftrightarrow P_i(t + \Delta t) - P_i(t) + \lambda P_i(t) \Delta t = \lambda P_{i-1}(t) \Delta t$$

$$\Leftrightarrow \lim_{\Delta t \rightarrow 0} \frac{P_i(t + \Delta t) - P_i(t)}{\Delta t} + \lambda P_i(t) = \lambda P_{i-1}(t)$$

$$\Rightarrow \frac{dP_i(t)}{dt} + \lambda P_i(t) = \lambda P_{i-1}(t) \quad P_i(0) = 0$$

$$P_i(t) = e^{-\lambda t} \frac{(\lambda t)^i}{i!}$$

Distributie

$$\varphi_I(i) = P_i(t) = e^{-\lambda t} \frac{(\lambda t)^i}{i!} \quad i = 0, 1, 2, \dots$$

$$\Phi_I(w) = \mathbf{P}(I \leq w) = e^{-\lambda t} \sum_{j \leq w} \frac{(\lambda t)^j}{j!}$$

$$e^x = \sum_{j=0}^{+\infty} \frac{x^j}{j!}$$

$$\Phi_I(+\infty) = e^{-\lambda t} \sum_{j=0}^{+\infty} \frac{(\lambda t)^j}{j!} = 1$$

Karakteristieken

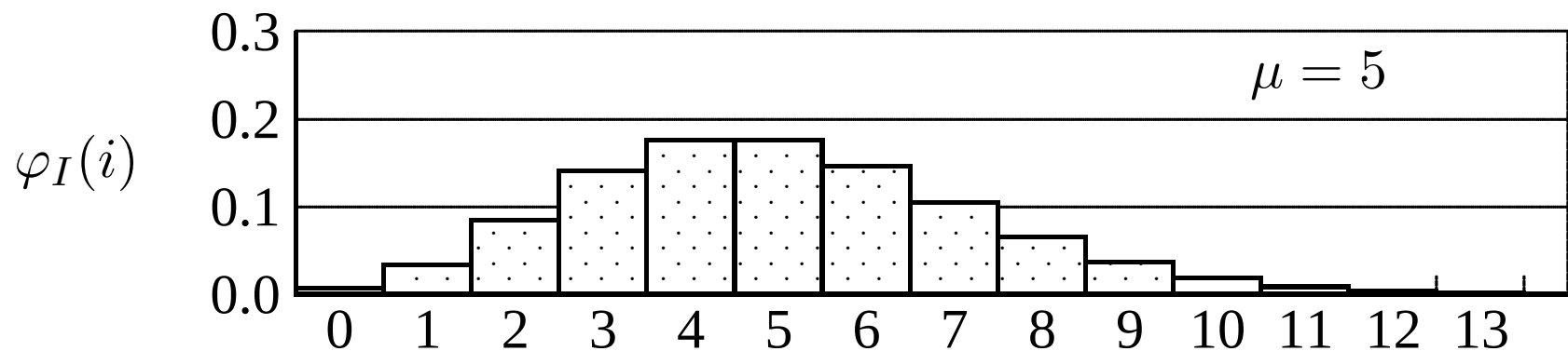
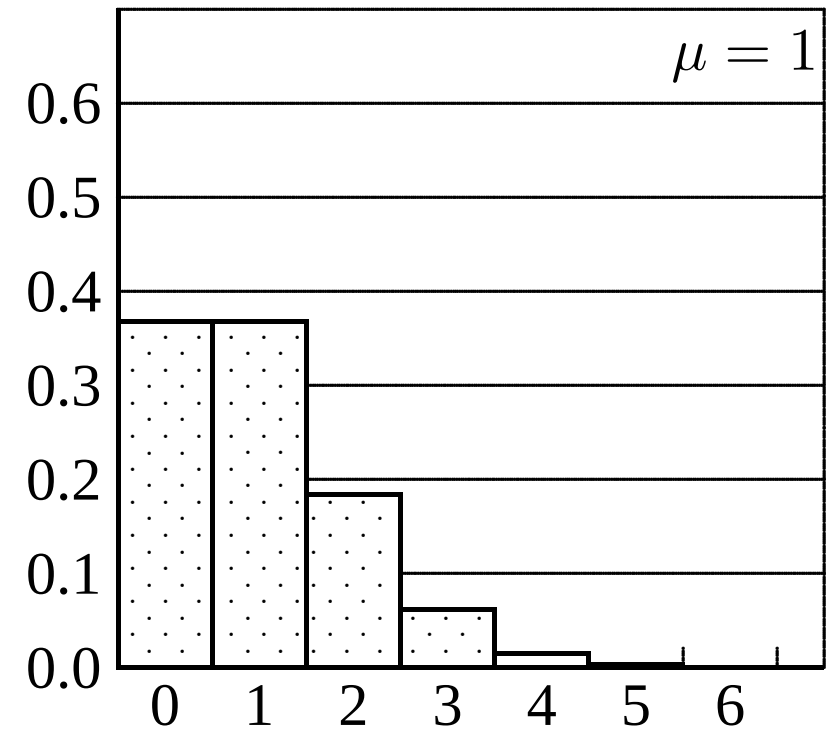
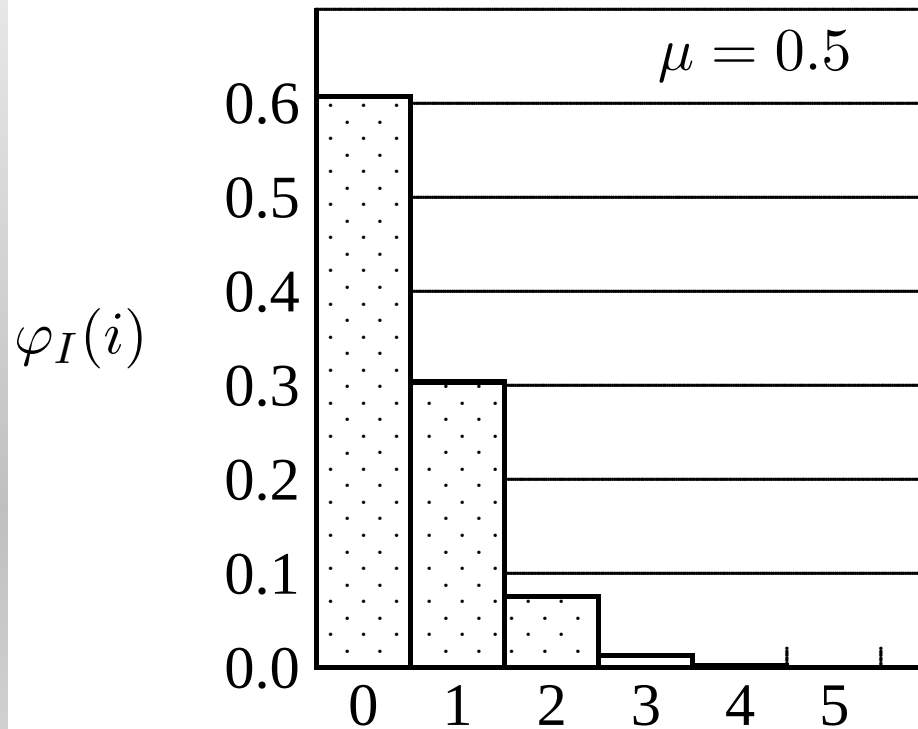
$$\mu_I = \sum_{i=0}^{\infty} i \varphi_I(i) = \lambda t$$

$$\sigma_I^2 = \sum_{i=0}^{\infty} i^2 \varphi_I(i) - \mu_I^2 = \lambda t$$

$$\varphi_I(i) = e^{-\mu} \frac{\mu^i}{i!} \quad i = 0, 1, 2, \dots$$

μ : parameter van de Poisson-distributie

Kansverdeling



Momentenfunctie

$$\begin{aligned}M_I(t) &= \sum_{i=0}^{\infty} e^{it} \varphi_I(i) \\&= \sum_{i=0}^{\infty} e^{it} e^{-\mu} \frac{\mu^i}{i!} \\&= e^{-\mu} \sum_{i=0}^{\infty} \frac{(e^t \mu)^i}{i!} \\&= e^{-\mu} e^{\mu e^t} \\&= e^{\mu(e^t - 1)}\end{aligned}$$

Momentenfunctie

Som van onafhankelijke Poisson verdeelde toevalsveranderlijken

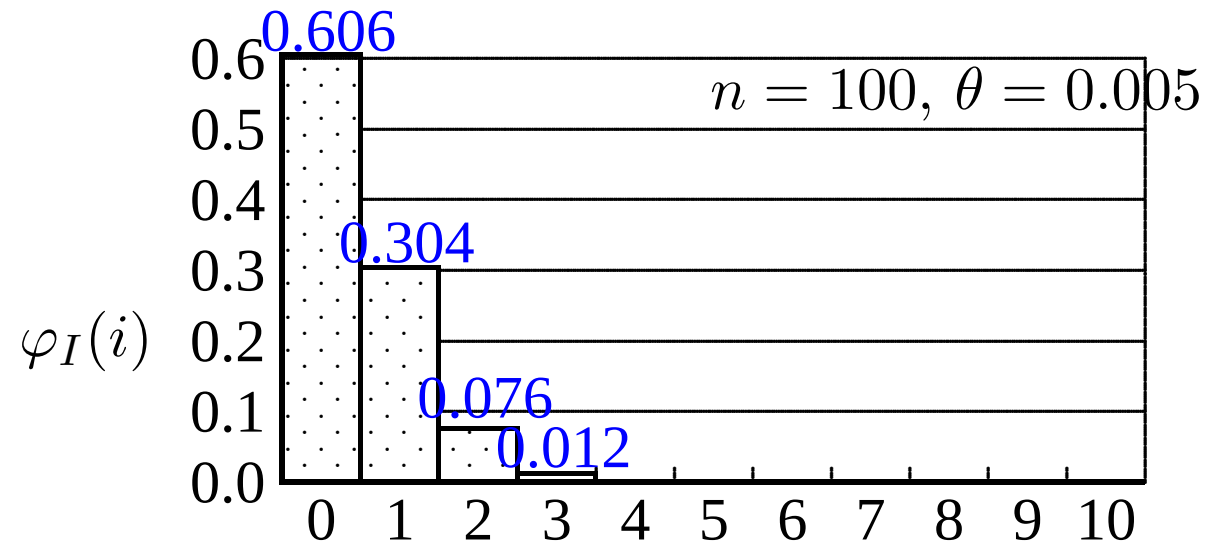
De som K van n onafhankelijke Poisson verdeelde toevalsveranderlijken I_j met parameter μ_j is Poisson verdeeld

met parameter $\sum_{j=1}^n \mu_j$.

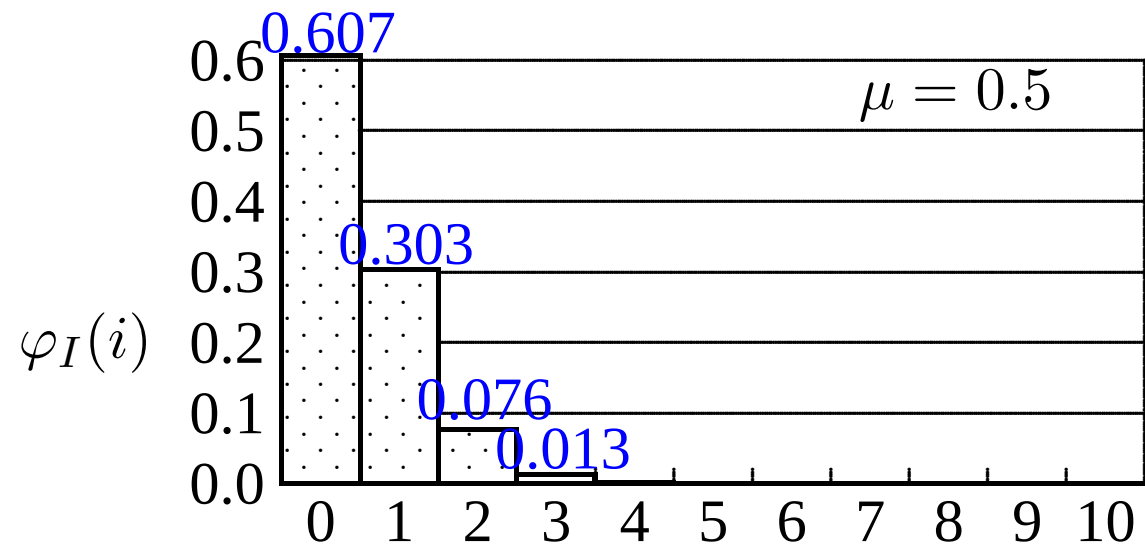
$$M_K(t) = \prod_{j=1}^n M_j(t) = \prod_{j=1}^n e^{\mu_j (e^t - 1)} = e^{(e^t - 1) \sum_{j=1}^n \mu_j}$$

Vergelijking

Binomiaal



Poisson



De Poisson-verdeling levert goede benaderingen voor de binomiale verdeling als n groot is en $\mu = n\theta$ klein.

Voorbeeld

Een radioactieve bron wordt geobserveerd gedurende vier verschillende tijdsintervallen van 6 seconden elk. Per seconde wordt er gemiddeld 0.5 deeltjes uitgezonden. Men neemt aan dat het aantal deeltjes dat uitgezonden wordt in het tijdsinterval $[0, t]$ verdeeld is volgens de Poisson-distributie.

- (i) $P(\text{in de vier intervallen} \geq 3 \text{ deeltjes uitgezonden})$
- (ii) $P(\text{in minstens 1 interval} \geq 3 \text{ deeltjes uitgezonden})$

Oplossing : I : het aantal uitgezonden deeltjes in $[0, t]$

$$\varphi_I(i) = e^{-\lambda t} \frac{(\lambda t)^i}{i!} \quad i = 0, 1, \dots \quad t = 6 \text{ seconden} \implies \lambda t = 3$$

$$P(I \geq 3) = 1 - P(I < 3) = 1 - e^{-\mu} \left(1 + \mu + \frac{\mu^2}{2!} \right) = 0.577$$

Voorbeeld

Een radioactieve bron wordt geobserveerd gedurende vier verschillende tijdsintervallen van 6 seconden elk. Per seconde wordt er gemiddeld 0.5 deeltjes uitgezonden. Men neemt aan dat het aantal deeltjes dat uitgezonden wordt in het tijdsinterval $[0, t]$ verdeeld is volgens de Poisson-distributie.

- (i) $P(\text{in de vier intervallen} \geq 3 \text{ deeltjes uitgezonden})$
- (ii) $P(\text{in minstens 1 interval} \geq 3 \text{ deeltjes uitgezonden})$

Oplossing : I : het aantal uitgezonden deeltjes in 6 seconden

$$P(I \geq 3) = 0.577$$

J : aantal intervallen met ≥ 3 uitgezonden deeltjes

J : binomiaal met $\theta_J = 0.577$

(i) $P(J = 4) = \theta^4 = 0.111$

(ii) $P(J \geq 1) = 1 - P(J = 0) = 1 - (1 - \theta)^4 = 0.968$

Veralgemeningen

De distributie van Poisson kan veralgemeend worden door het constant zijn van λ te laten varen.

Als voorbeeld kunnen we het aantal slachtoffers van een besmettelijke ziekte gedurende een tijd t beschouwen. De besmettingsparameter neemt toe als het aantal nieuwe zieken in eenzelfde tijdsduur toeneemt. Daarentegen neemt af als de tijdsduur voor evenveel nieuwe zieken toeneemt.

Een van de veralgemeningen is de distributie van Polya.

Poisson en binomiaal

Bepaal $\varphi_J(j)$ als J telt hoeveel verschijnselen uit een Poisson proces (met parameter μ) ook voldoen aan een zekere eigenschap (met constante kans θ).

- een Poisson proces $I : \varphi_I(i) = e^{-\mu} \frac{\mu^i}{i!} \quad i = 0, 1, \dots$
- binomiaal proces met parameter θ

$$\begin{aligned}\varphi_J(j) &= \mathbf{P}(J = j) = \mathbf{P}((J = j) \cdot (\sum_{i=0}^{\infty} (I = i))) \quad j = 0, 1, \dots \\&= \sum_{i=0}^{\infty} \mathbf{P}((J = j) \cdot (I = i)) = \sum_{i=0}^{\infty} \mathbf{P}(I = i) \mathbf{P}(J = j \mid I = i) \\&= \sum_{i=j}^{\infty} \mathbf{P}(I = i) \mathbf{P}(J = j \mid I = i) \text{ want } P(J > I) = 0 \\&= \sum_{i=j}^{\infty} e^{-\mu} \frac{\mu^i}{i!} \binom{i}{j} \theta^j (1 - \theta)^{i-j}\end{aligned}$$

Poisson en binomiaal

Bepaal $\varphi_J(j)$ als J telt hoeveel verschijnselen uit een Poisson proces (met parameter μ) ook voldoen aan een zekere eigenschap (met constante kans θ).

$$\begin{aligned}\varphi_J(j) &= \sum_{i=j}^{\infty} e^{-\mu} \frac{\mu^i}{i!} \binom{i}{j} \theta^j (1-\theta)^{i-j} \\&= e^{-\mu} \sum_{i=j}^{\infty} \frac{\mu^{j+(i-j)}}{i!} \frac{i!}{(i-j)! j!} \theta^j (1-\theta)^{i-j} \\&= \frac{e^{-\mu} (\theta \mu)^j}{j!} \sum_{i-j=0}^{\infty} \frac{1}{(i-j)!} (\mu (1-\theta))^{i-j} \\&= \frac{e^{-\mu} (\theta \mu)^j}{j!} \sum_{k=0}^{\infty} \frac{1}{k!} (\mu (1-\theta))^k \\&= \frac{e^{-\mu} (\theta \mu)^j}{j!} e^{\mu(1-\theta)} = e^{-\mu\theta} \frac{(\theta \mu)^j}{j!}\end{aligned}$$

J : Poisson $\mu_J = \mu \theta$

De discrete uniforme verdeling

Een uniform verdeelde discrete veranderlijke X is een veranderlijke waarbij de kans op het voorkomen van een enkelvoudige gebeurtenis uit de populatie voor elke enkelvoudige gebeurtenis dezelfde is.

Distributie – karakteristieken

waarden $X : x_1 < x_2 < \dots < x_m$

$$\varphi_X(x_i) = \mathbf{P}(X = x_i) = \frac{1}{m} \quad i = 1, 2, \dots, m$$

$$\Phi_X(w) = \mathbf{P}(X \leq w) = \sum_{x_i \leq w} \frac{1}{m}$$

$$\Phi_X(x_i) = \mathbf{P}(X \leq x_i) = \frac{i}{m} \quad i = 1, 2, \dots, m$$

$$\mu_X = \frac{1}{m} \sum_{i=1}^m x_i$$

$$\sigma_X^2 = \frac{1}{m} \sum_{i=1}^m x_i^2 - \mu_X^2$$

Bijzonder geval

waarden $I : 1 < 2 < \dots < m$

$$\mu_I = \frac{m+1}{2}$$

$$\sigma_I^2 = \frac{m^2 - 1}{12}$$