BAB 2

Elemen Probabilitas

2.1 Ruang Sampel dan Peristiwa

Perhatikan sebuah eksperimen yang hasilnya belum diketahui di muka. S, yang disebut ruang sampel eksperimen, menunjukkan kumpulan semua hasil yang mungkin dicapai. Sebagai contoh, jika eksperimen tersebut berupa pacuan di antara tujuh ekor kuda yang diberi nomor 1 sampai 7, maka

$$S = \{\text{semua urutan } (1, 2, 3, 4, 5, 6, 7)\}$$

Sebagai contoh, hasilnya (3, 4, 1, 7, 6, 5, 2) berarti bahwa kuda nomor 3 tiba lebih dahulu, kuda nomor 4 kedua, dan seterusnya.

Himpunan bagian A dari ruang sampel dikenal sebagai suatu peristiwa. Dengan kata lain, peristiwa adalah suatu himpunan yang terdiri atas hasil yang mungkin diperoleh dalam eksperimen. Jika hasil eksperimen ada di dalam A, kita katakan bahwa A telah terjadi. Sebagai contoh, di atas jika

$$A = \{\text{semua hasil dalam } S \text{ yang dimulai dengan } 5\}$$

maka A adalah peristiwa di mana kuda nomor 5 masuk lebih dahulu.

Untuk dua peristiwa A dan B kita tetapkan peristiwa baru $A \cup B$, yang disebut gabungan *(union)* A dan B, yang terdiri atas semua hasil yang ada di A atau B atau di A maupun B. Begitu pula halnya, kita menetapkan peristiwa, AB, yang disebut irisan *(intersection)* A dan B, terdiri atas semua hasil yang ada baik di A maupun B. "Dengan kata lain, peristiwa $A \cup B$ terjadi jika A atau B terjadi; sementara peristiwa AB terjadi jika baik A maupun B terjadi. Kita juga dapat menetapkan gabungan dan irisan lebih dari dua peristiwa. Secara khusus, gabungan peristiwa $A_1, \ldots, A_n$—yang ditandai dengan $\cup_{i=1}^{n} A_i$—ditetapkan untuk

$\cup_{i=1}^{n} A_i$

terdiri atas semua hasil yang setidaknya ada pada salah satu dari A_i. Begitu pula, irisan peristiwa $A_1, \ldots, A_n$ — yang ditandai dengan $A_1A_2 \ldots A_n$ — ditetapkan untuk terdiri atas semua hasil yang ada di semua A_i.

Untuk setiap peristiwa A kita definisikan peristiwa A^c, yang dirujuk sebagai komplemen A, untuk terdiri atas semua hasil dalam ruang sampel S yang tidak ada dalam A. Dengan kata lain, A^c terjadi hanya jika A tidak terjadi. Karena hasil eksperimen harus terletak dalam ruang sampel S, maka S^c tidak mengandung hasil apa pun dan dengan demikian tidak mungkin terjadi. Kita sebut S^c himpunan nol dan menandainya dengan $\varnothing$. Jika $AB = \varnothing$ maka A dan B tidak dapat sama-sama terjadi (karena tidak ada hasil yang ada baik dalam A maupun B), kita katakan bahwa A dan B berdiri sendiri-sendiri.

2.2. Aksioma Probabilitas

Andaikan bahwa untuk setiap peristiwa A suatu eksperimen yang memiliki ruang sampel S ada sebuah bilangan, yang ditandai dengan $P(A)$ dan disebut probabilitas peristiwa A, yang sesuai dengan tiga aksioma berikut ini:

Aksioma 1:

$$0 \leq P(\mathrm{A}) \leq 1$$

Aksioma 2:

$$P(S) = 1$$

Aksioma 3:
Untuk tiap urutan peristiwa yang berdiri sendiri $A_1, A_2, \ldots$

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i), \qquad n = 1, 2, \ldots, \infty$$

Jadi Aksioma 1 menyatakan probabilitas bahwa hasil dari eksperimen terletak di dalam A adalah angka di antara 0 dan 1. Aksioma 2 menyatakan bahwa dengan probabilitas 1 hasil ini merupakan anggota dari ruang sampel; dan Aksioma 3 menyatakan bahwa untuk himpunan peristiwa yang berdiri sendiri mana pun, probabilitas bahwa setidaknya salah satu dari peristiwa ini terjadi sama dengan jumlah probabilitas mereka masing-masing.

Ketiga aksioma ini dapat digunakan untuk membuktikan bermacam hasil tentang probabilitas. Sebagai contoh, karena A dan A^c selalu berdiri sendiri-sendiri, dan karena $A \cup A^c = S$, kita mendapatkan dari Aksioma 2 dan 3 bahwa

$$1 = P(S) = P(A \cup A^c) = P(A) + P(A^c)$$

atau sama dengan

$$P(A^c) = 1 - P(A)$$

Dalam kata-kata, probabilitas bahwa suatu peristiwa tidak terjadi adalah 1 minus probabilitas bahwa peristiwa itu terjadi.

2.3 Probabilitas bersyarat dan Independensi

Pertimbangkan sebuah eksperimen berupa lemparan koin sebanyak dua kali, dengan setiap kali mencatat apakah hasilnya gambar (H) atau angka (T). Ruang sampel dari eksperimen ini dapat dianggap sebagai himpunan empat hasil berikut ini:

$$S = \{(\mathrm{H}, \mathrm{H}), (\mathrm{H}, \mathrm{T}), (\mathrm{T}, \mathrm{H}), (\mathrm{T}, \mathrm{T})\}$$

di mana, sebagai contoh, (H, T) berarti bahwa lemparan pertama koin menghasilkan gambar (H) dan lemparan kedua angka (T). Andaikan sekarang bahwa masing dari keempat hasil yang dimungkinkan sama mungkinnya untuk terjadi, dan dengan begitu probabilitasnya $\frac{1}{4}$. Andaikan lebih lanjut kita mengamati bahwa lemparan pertama menghasilkan sisi gambar. Lalu, dengan informasi ini, bagaimana probabilitas bahwa kedua lemparan akan menghasilkan sisi gambar? Untuk menghitung probabilitas ini kita mengajukan penalaran berikut: dengan lemparan awal menghasilkan sisi gambar, maka paling banyak mungkin ada dua hasil dari eksperimen kita, yaitu, (H, H) atau (H, T). Selain itu, ketika masing-masing hasil ini semula memiliki probabilitas yang sama untuk terjadi, maka hasil-hasil tersebut seharusnya tetap memiliki probabilitas yang sama. Dengan kata lain, dengan lemparan pertama yang menghasilkan sisi gambar, probabilitas (bersyarat) dari masing-masing hasil (H, H) dan (H, T) adalah $\frac{1}{2}$, sementara probabilitas (bersyarat) dua hasil lainnya adalah 0. Oleh karena itu, probabilitas yang diinginkan adalah $\frac{1}{2}$.

Seandainya A dan B secara berturut-turut menandai peristiwa-peristiwa di mana kedua lemparan menghasilkan sisi gambar dan peristiwa di mana lemparan pertama menghasilkan sisi gambar, maka probabilitas yang diperoleh di atas disebut probabilitas bersyarat A dengan B telah terjadi dan ditandai dengan

$$P(A \mid \mathrm{B})$$

Rumus umum untuk $P(A \mid \mathrm{B})$ yang absah untuk semua eksperimen dan peristiwa A dan B dapat diperoleh dengan cara yang sama seperti yang diberikan sebelumnya. Caranya adalah, jika peristiwa B terjadi, maka agar A terjadi maka peristiwa aktual perlu ada pada A maupun B; yaitu, harus ada dalam AB. Karena kita tahu bahwa B sudah terjadi, maka B menjadi ruang sampel baru kita dan karenanya probabilitas bahan peristiwa AB terjadi akan sama dengan probabilitas AB relatif dengan probabilitas B. Artinya,

$$P(A \mid B) = \frac{P(AB)}{P(B)}$$

Seperti ditunjukkan oleh contoh lemparan koin, $\mathrm{P}(A \mid B)$, probabilitas bersyarat A, dengan B telah terjadi, umumnya tidak sama dengan $P(A)$, probabilitas tak bersyarat A. Dengan kata lain, mengetahui bahwa B telah terjadi umumnya mengubah probabilitas bahwa A terjadi (bagaimana jika keduanya berdiri sendiri-sendiri?). Dalam kasus khusus di mana $P(A \mid B)$ sama dengan $P(A)$, kita katakan bahwa A dan B independen (bebas). Karena $P(A \mid B) = P(AB)/P(B)$, kita lihat bahwa A tidak bergantung pada B jika

$$P(AB) = P(A)P(B)$$

Karena hubungan ini simetris pada A dan B, maka setiap kali A tidak bergantung pada B, B juga tidak bergantung pada A.

2.4. Variabel Acak

Ketika suatu eksperimen dilakukan, kita kadang terutama khawatir akan nilai dari sejumlah kuantitas numerik yang ditentukan oleh hasilnya. Kuantitas yang ditentukan oleh hasil eksperimen ini dikenal sebagai variabel acak.

Fungsi distribusi kumlatif, atau lebih sederhana fungsi distribusi, F, dari variabel acak X ditetapkan untuk bilangan nyata x mana pun dengan

$$F(x) = P\{X \le x\}$$

Variabel acak yang dapat menggunakan bilangan terbatas *(finite number)* atau bilangan terhitung *(countable number)* dari nilai yang dimungkinkan dikatakan diskret. Untuk variabel acak diskret X kita tetapkan fungsi masa probabilitasnya $p(x)$ dengan

$$p(x) = P\{X = x\}$$

Jika X adalah variabel acak diskret yang menggunakan salah satu dari nilai yang mungkin $x_1, x_2, \ldots$, maka, karena X harus menggunakan salah satu dari nilai ini, kita mendapatkan

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

Contoh 2a andaikan bahwa X menggunakan salah satu dari nilai 1, 2, atau 3. Jika

$$p(1) = \tfrac{1}{4}, \quad p(2) = \tfrac{1}{3}$$

maka, karena $p(1) + p(2) + p(3) = 1$, jadi $p(3) = \frac{5}{12}$

Sementara variabel acak diskret menerima himpunan terhitung dari nilai yang mungkin, kita sering harus memperhatikan variabel acak yang himpunan nilai yang mungkinnya berupa sebuah interval. Kita katakan bahwa variabel acak X adalah variabel acak kontinu jika ada fungsi nonegatif $f(x)$ yang ditetapkan untuk semua bilangan nyata x dan memiliki sifat yang untuk himpunan C mana pun dari bilangan nyata

$$P\{X \in C\} = \int_C f(x)\,dx \tag{2.1}$$

Fungsi f disebut fungsi densitas probabilitas dari variabel acak X.

Hubungan antara distribusi kumulatif $f(\cdot)$ dan densitas probabilitas $f(\cdot)$ dinyatakan dengan

$$F(a) = P\{X \in (-\infty, a]\} = \int_{-\infty}^{a} f(x)\,dx$$

Diferensiasi kedua sisi menghasilkan

$$\frac{d}{da}F(a) = f(a)$$

Dengan kata lain, densitas adalah derivatif dari fungsi distribusi kumulatif. Penafsiran yang agak lebih intuitif mengenai fungsi densitas dapat diperoleh dari Persamaan (2.1) sebagai berikut:

$$P\left\{a - \frac{\varepsilon}{2} \leq X a + \frac{\varepsilon}{2}\right\} = \int_{a-\varepsilon/2}^{a+\varepsilon/2} f(x)\,dx \approx \varepsilon f(a)$$

jika ε kecil. Dengan kata lain, probabilitas bahwa X akan ada di dalam interval dengan kepanjangan ε sekitar titik a kira-kira adalah $\varepsilon f(a)$. Dari sini, kita lihat bahwa $f(a)$ adalah ukuran dari seberapa mungkin bahwa variabel acak akan mendekati a.

Dalam banyak eksperimen kita tertarik tidak hanya akan fungsi distribusi probabilitas variabel acak individual, tetapi juga akan hubungan di antara dua fungsi atau lebih. Untuk menetapkan hubungan di antara dua variabel acak, kita tetapkan fungsi distribusi probabilitas kumulatif bersama X dan Y dengan

$$F(x, y) = P\{X \leq x,\ Y \leq y\}$$

Jadi $F(x, y)$ menetapkan probabilitas bahwa X kurang dari atau sama dengan x dan secara serentak Y kurang dari atau sama dengan y.

Jika X dan Y sama-sama variabel acak diskret, maka kita tetapkan fungsi massa probabilitas bersama X dan Y dengan

$$p(x, y) = P\{X = x, Y = y\}$$

Demikian pula halnya, kita katakan bahwa X dan Y kontinu secara bersama-sama, dengan fungsi densitas probabilitas bersama $f(x, y)$ jika untuk himpunan apa pun dari bilangan nyata C dan D

$$P\{X \in C, Y \in D\} = \iint\limits_{\substack{x \in C \\ y \in D}} f(x, y)\, dx\, dy$$

Variabel acak X dan Y dikatakan bebas atau independen jika untuk dua himpunan apa pun dari bilangan nyata C dan D

$$P\{X \in C, Y \in D\} = P\{X \in C\}\, P\{Y \in D\}$$

Jadi X dan Y independen jika untuk semua himpunan C dan D, peristiwa $A = \{X \in C\}$, $B = \{Y \in D\}$ independen. Secara umum, X dan Y independen jika pengetahuan tentang nilai salah satu dari mereka tidak mempengaruhi distribusi probabilitas yang lain. Variabel acak yang tidak independen dikatakan independen.

Dengan menggunakan ketiga aksioma probabilitas, kita dapat memperlihatkan bahwa variabel acak X dan Y akan independen hanya jika

$$P\{X \leq x, Y \leq y\} = P\{X \leq x\}\, P\{Y \leq y\}, \quad \text{untuk semua } x, y$$

Yaitu, X dan Y akan independen hanya jika

$$F(x, y) = F_x(x)\, F_y(y), \quad \text{untuk semua } x, y$$

di mana F_X dan F_Y merupakan fungsi distribusi dari X dan Y.

2.5 Harapan

Salah satu konsep paling berguna dalam probabilitas adalah konsep harapan variabel acak. Jika X adalah variabel acak diskret yang menggunakan salah satu nilai yang mungkin $x_1, x_2, \ldots$, maka *harapan* atau *nilai yang diharapkan* dari X, yang juga disebut *rerata* X dan ditandai dengan $E[X]$, ditetapkan dengan

$$E[X] = \sum_i x_i \, P\{X = x_i\} \tag{2.2}$$

Dalam kata-kata, nilai yang diharapkan dari X merupakan rata-rata tertimbang dari nilai yang mungkin yang dapat digunakan oleh X, tiap nilai ditimbang dengan probabilitas yang diambil oleh X. Sebagai contoh, jika fungsi massa probabilitas X diberikan dengan

$$p(0) = \tfrac{1}{2} = p(1)$$

maka

$$E[X] = 0\,(\tfrac{1}{2}) + 1\,(\tfrac{1}{2}) = (\tfrac{1}{2})$$

hanyalah rata-rata biasa dari dua nilai yang mungkin 0 dan 1 yang dapat diambil oleh X. Sebaliknya, jika

$$p(0) = \tfrac{1}{3}, \qquad p(1) = \tfrac{2}{3}$$

maka

$$E[X] = 0\,(\tfrac{2}{3}) + 1\,(\tfrac{2}{3}) = \tfrac{2}{3}$$

merupakan rata-rata tertimbang dari dua nilai yang mungkin 0 dan 1 di mana nilai 1 diberikan bobot dua kali lebih besar daripada nilai 0 karena $p(1) = 2p(0)$.

Contoh 2b Jika I merupakan variabel acak indikator untuk peristiwa A, yaitu, jika

$$I = \begin{cases} 1 & \text{Jika } A \textit{ terjadi} \\ 0 & \text{Jika } A \textit{ tidak terjadi} \end{cases}$$

maka

$$E[I] = 1P(A) + 0P(A^c) = P(A)$$

Oleh karena itu harapan variabel acak indikator untuk peristiwa A hanyalah probabilitas bahwa A terjadi.

Jika X merupakan variabel acak kontinu yang memiliki fungsi densitas f, maka, analog dengan Persamaan (2.2), kita mendefinisikan nilai X yang diharapkan dengan

$$E[X] = \int_{-\infty}^{\infty} xf(x)\,dx$$

Contoh 2c jika fungsi densitas probabilitas X diberikan dengan

$$f(x) = \begin{cases} 3x^2 & \text{jika } 0 < x < 1 \\ 0 & \text{lainnya} \end{cases}$$

maka

$$E[X] = \int_0^1 3x^3\,dx = \tfrac{3}{4}$$

Andaikan saja bahwa kita ingin menentukan nilai yang diharapkan bukan dari variabel acak X tetapi dari variabel acak $g(X)$, di mana g merupakan fungsi yang diberikan. Karena $g(X)$ menggunakan nilai $g(x)$ ketika X menggunakan nilai x, tampaknya intuitif bahwa $E[g(X)]$ harus berupa rata-rata tertimbang dari nilai yang mungkin $g(x)$, bobot yang diberikan pada $g(x)$ yang sama dengan probatilias (atau densitas probabilitas dalam kasus kontinu) bahwa X akan sama dengan x. Sebenarnya, yang sebelumnya dapat dibuktikan benar dan dengan begitu kita memperoleh hasil berikut.

Proposisi:

Jika X merupakan variabel acak diskret yang memiliki fungsi massa probabilitas $p(x)$, maka

$$E[g(X)] = \sum_x g(x)\,p(x)$$

sementara jika X kontinu dengan fungsi densitas probabilitas $f(x)$, maka

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)\, f(x)\, dx$$

Konsekuensi dari proposisi di atas adalah sebagai berikut.

Konsekuensi wajar:

Jika a dan b konstan, maka

$$E[aX + b] = aE[X] + b$$

Bukti. Dalam kasus diskret

$$E[aX + b] = \sum_{x} (ax + b)p(x)$$

$$= a\sum_{x} xp(x) + b\sum_{x} p(x)$$

$$= aE[X] + b$$

Karena bukti dalam kasus kontinu serupa, hasilnya pun ditetapkan.

Dapat dilihat di sini bahwa harapan merupakan operasi linear dalam pengertian bahwa untuk dua variabel acak X_1 dan X_2 mana pun

$$E[X_1 + X_2] = E[X_1] + E[X_2]$$

yang dengan mudah digeneralisasikan untuk menghasilkan

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

2.6 Variansi

Mengingat $E[X]$, nilai yang diharapkan dari variabel acak X, merupakan rata-rata tertimbang dari nilai yang mungkin dari X, maka tidak ada informasi yang dihasilkan tentang variasi dari nilai-nilai ini. Salah satu cara untuk mengukur variasi ini adalah dengan mempertimbangkan nilai rata-rata pangkat dari perbedaan antara X dan $E[X]$. Dengan demikian kita mendapatkan definisi berikut ini.

Definisi:

Jika X merupakan variabel acak dengan rerata μ, maka variansi X, yang ditandai dengan $Var(X)$, ditetapkan dengan

$$\text{Var}(X) = E[(X - \mu)^2]$$

Rumus alternatif untuk Var(X) diperoleh sebagai berikut:

$$\begin{aligned}\text{Var}(X) &= E[(X - \mu)^2] \\ &= E[X^2 - 2\mu X + \mu^2] \\ &= E[X^2] - E[2\mu X] + E[\mu^2] \\ &= E[X^2] - 2\mu E[X] + \mu^2 \\ &= E[X^2] - \mu^2\end{aligned}$$

Yaitu,

$$\text{Var}(X) = E[X^2] - (E[X])^2$$

Identitas yang berguna, yang buktinya dibiarkan sebagai latihan, adalah untuk konstan a dan b mana pun

$$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$$

Mengingat nilai yang diharapkan dari jumlah variabel acak sama dengan jumlah harapan, hasil yang sesuai pada umumnya tidak berlaku untuk variansi. Akan tetapi, hasil ini berlaku dalam kasus khusus yang penting di mana variabel acak independen. Sebelum membuktikan ini, marilah kita definisikan konsep kovarian di antara variabel acak.

Definisi:

Kovarian dari dua variabel acak X dan Y, yang ditandai Cov(X, Y), ditetapkan dengan

$$\mathrm{Cov}(X, Y) = E[\,(X - \mu_x)\,(Y - \mu_y)]$$

di mana $\mu_x = E[X]$ dan $\mu_y = E[Y]$.

Ekspresi yang berguna untuk Cov(X, Y) diperoleh dengan memperluas sisi kanan dari persamaan di atas dan kemudian menggunakan linearitas harapannya. Ini menghasilkan

$$\begin{aligned}\mathrm{Cov}\,(X, Y) &= E\,[XY - \mu_x Y - X\mu_y + \mu_x\mu_y] \\ &= E\,[XY] - \mu_x E\,[Y] - E\,[X]\,\mu_y + \mu_x\mu_y \\ &= E\,[XY] - E\,[X]\,E\,[Y] \qquad (2.3)\end{aligned}$$

Kita sekarang mengambil ekspresi untuk Var($X + Y$) berkaitan dengan variansi individual mereka dan kovarian di antara mereka. Karena

$$E[X + Y] = E[X] + E[Y] = \mu_x + \mu_y$$

kita lihat bahwa

$$\begin{aligned}Var\,(X + Y) &= E\,[(X + Y - \mu_x - \mu_y)^2] \\ &= E\,[(X - \mu_x)^2 + (Y - \mu_y)^2 + 2\,(X - \mu_x)\,(Y - \mu_y)] \\ &= E\,[(X - \mu_x)^2] + E\,[(Y - \mu_y)^2] + 2E\,[(X - \mu_x)\,(Y - \mu_y)] \\ &= Var\,(X) + Var\,(Y) + 2\,Cov\,(X, Y) \qquad (2.4)\end{aligned}$$

Kita mengakhiri bagian ini dengan memperlihatkan bahwa variansi dari jumlah variabel acak independen sama dengan jumlah variansi mereka.

Proposisi:

Jika X dan Y merupakan variabel acak independen maka

$$\mathrm{Cov}(X, Y) = 0$$

dan dengan demikian, dari Persamaan (2.4),

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

Bukti. Dari Persamaan (2.3) diketahui bahwa kita perlu memperlihatkan bahwa $E[XY] = E[X]E\{Y]$. Sekarang dalam kasus diskret,

$$\begin{aligned} E[XY] &= \sum_j \sum_i x_i y_j P\{X = x_i,\ Y = y_j\} \\ &= \sum_j \sum_i x_i y_j P\{X = x_i\} P\{Y = y_j\} \qquad \text{by independence} \\ &= \sum_j y_j P\{Y = y_j\} \sum_i x_i P\{X = x_i\} \\ &= E[Y]E[X] \end{aligned}$$

Karena argumen serupa berlaku dalam kasus kontinu, hasilnya pun terbukti.

2.7. Ketidaksamaan Chebyshev dan Kaidah Bilangan Besar

Kita mulai dengan hasil yang dikenal sebagai ketidaksamaan*(inequality)* Markov.

Proposisi: Ketidaksamaan Markov

Jika X menggunakan hanya nilai nonnegatif, maka untuk nilai a > 0 mana pun

$$P\{X \geq a\} \leq \frac{E[X]}{a}$$

Bukti. Kita memberikan bukti untuk kasus di mana X merupakan variabel acak kontinu yang memiliki fungsi densitas f

$$\begin{aligned}
E[X] &= \int_0^\infty x f(x)\, dx \\
&= \int_0^a x f(x)\, dx + \int_a^\infty x f(x)\, dx \\
&\geq \int_a^\infty x f(x)\, dx \\
&\geq \int_a^\infty a f(x)\, dx \qquad \text{since } x f(x) \geq a f(x) \text{ when } x \geq a \\
&= a \int_a^\infty f(x)\, dx = a P\{X \geq a\}
\end{aligned}$$

dan hasilnya terbukti.

Sebagai konsekuensi wajar kita memiliki apa yang disebut ketidaksamaan Chebyshev, yang menyatakan bahwa probabilitas variabel acak berbeda dengan *rerata*-nya dengan lebih dari k simpangan bakunya (deviasi standar) dibatasi dengan $1/k^2$, di mana simpangan baku variabel acaknya ditentukan dengan akar pangkat dua variansinya.

Konsekuensi wajar: Ketidaksamaan Chebyshev

Jika X adalah variabel acak yang memiliki rerata μ *dan variansi* σ

$$P\{|X - \mu| \geq k\sigma\} \leq \frac{1}{k^2}$$

Bukti. Karena $(X - \mu)^2/\sigma$ *merupakan variabel acak nonnegatif yang* rerata-*nya adalah*

$$E\left[\frac{(X-\mu)^2}{\sigma^2}\right] = \frac{E[(X-\mu)^2]}{\sigma^2} = 1$$

kita mendapatkan dari ketidaksamaan Markov bahwa

$$P\left\{\frac{(X-\mu)^2}{\sigma^2} \geq k^2\right\} \leq \frac{1}{k^2}$$

Hasilnya sekarang karena ketidaksamaan $(X - \mu)^2/\sigma^2 \geq k^2$ adalah sama dengan ketidaksamaan $|X - \mu| \geq k\sigma$

Kita sekarang menggunakan ketidaksamaan Chebyshev untuk membuktikan kaidah yang lemah dari bilangan besar, yang menyatakan bahwa probabilitas rata-rata n persyaratan urutan pertama dari variabel acak yang independen dan identik berbeda dengan rerata-nya dengan lebih dari menjadi 0 ketika n menjadi tak terhingga.

Dalil: Kaidah Lemah Bilangan Besar

Misalkan $X_1, X_2, \ldots$ merupakan urutan variabel acak yang independen dan terdistribusi secara identik yang memiliki rerata μ Maka untuk tiap $\varepsilon > 0$.

$$P\left\{\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| > \epsilon\right\} \to 0 \qquad \text{as } n \to \infty$$

Bukti. Kita memberikan bukti di bawah asumsi tambahan bahwa variabel acak X_i memiliki variansi terbatas σ^2. Sekarang

$$E\left[\frac{X_1 + \cdots + X_n}{n}\right] = \frac{1}{n}(E[X_1] + \cdots + E[X_n]) = \mu$$

dan

$$\operatorname{Var}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n^2}[\operatorname{Var}(X_1) + \cdots + \operatorname{Var}(X_n)] = \frac{\sigma^2}{n}$$

di mana pertanyaan di atas menggunakan fakta bahwa variansi jumlah variabel acak independen sama dengan jumlah variansi mereka. Oleh karena itu, dari ketidaksamaan Chebyshev, didapatkan bahwa untuk tiap k positif

$$P\left\{\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \frac{k\sigma}{\sqrt{n}}\right\} \leq \frac{1}{k^2}$$

Oleh karena itu, untuk tiap $\varepsilon > 0$, dengan memisalkan k menjadi $k\sigma/\sqrt{n} = \varepsilon$, yaitu, dengan memisalkan $k^2 = n\varepsilon^2/\sigma^2$, kita melihat bahwa

$$P\left\{\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right\} \leq \frac{\sigma^2}{n\epsilon^2}$$

yang menetapkan hasilnya.

Generalisasi kaidah lemah adalah kaidah kuat bilangan besar, yang menyatakan bahwa, dengan probabilitas 1,

$$\lim_{n \to \infty} \frac{X_1 + \ldots + X_n}{n} = \mu$$

Yaitu, dengan kepastian, rata-rata jangka panjang urutan variabel acak independen dan yang terdistribusi secara identik akan bertemu pada rerata-nya.

2.8 Beberapa Variabel Acak Diskret

Ada jenis-jenis tertentu variabel acak yang sering muncul dalam aplikasi. Dalam bagian ini, kita membahas beberapa variabel acak diskret.

Variabel Acak Binomial

Andaikan bahwa n percobaan independen, yang masing-masing "berhasil" dengan probabilitas p, harus dijalankan. Jika X menggambarkan jumlah akses yang terjadi dalam n percobaan, maka X dikatakan merupakan variabel acak binomial dengan parameter (n, p). Fungsi massa probabilitasnya diberikan dengan

$$p_i \equiv P\{X = i\} = \binom{n}{i} p^i(1-p)^{n-i}, \qquad i = 0, 1, \ldots, n \tag{2.5}$$

di mana $\binom{n}{i}$ = n!/[n - i)!i!] adalah koefisien binomial, sama dengan jumlah himpunan bagian yang berbeda dari elemen i yang dapat dipilih dari himpunan elemen n.

Keabsahan Persamaan (2.5) dapat dilihat dengan lebih dahulu memperhatikan bahwa probabilitas urutan tertentu hasil mana pun yang menghasilkan i keberhasilan dan n - i kegagalan adalah, dengan independensi percobaan yang diasumsikan, $p^i(1\text{-}p)^{n-i}$. Persamaan yang menghasilkan (2.5) keberhasilan $\binom{n}{i}$ yang dapat dilihat dengan memperhatikan bahwa ada $\binom{n}{i}$ pilihan yang berbeda dari i percobaan yang menghasilkan keberhasilan.

Variabel acak binomial (1, p) disebut variabel acak Bernoulli. Karena suatu variabel acak X binomial (n,p) merupakan jumlah keberhasilan dalam n percobaan independen, yang masing-masing menghasilkan keberhasilan dengan probabilitas p, kita dapat merepresentasikan sebagai berikut:

$$X = \sum_{i=1}^{n} X_i \tag{2.6}$$

di mana

$$X_i = \begin{cases} 1 & \text{Jika percobaan ke} - i \text{ merupakan keberhasilan} \\ 0 & \text{lainnya} \end{cases}$$

Sekarang

$$\begin{aligned} E[X_i] &= P\{X_i = 1\} = p \\ Var(X_i) &= E[X_i^2] - E([X_i])^2 \\ &= p - p^2 = p(1-p) \end{aligned}$$

di mana persamaan di atas menggunakan fakta bahwa $X_{i2} = X_i$ (karena $0^2 = 0$ dan $1^2 = 1$). Jadi representasi (2.6) menghasilkan bahwa, untuk suatu variabel acak X binomial (n,p),

$$E[X] = \sum_{i=1}^{n} E[X_i] = np$$

$$Var(X) = \sum_{i=1}^{n} Var(X_i) \qquad \text{karena} X_i \text{ independen}$$

$$= np(1-P)$$

Rumus rekursif berikut ini yang mengekspresikan p_{i+1} dalam hal p_i berguna ketika penghitungan probabilitas binomial:

$$p_{i+1} = \frac{n!}{(n-i-1)!\,(i+1)!} p^{i+1} (1-p)^{n-i-1}$$

$$= \frac{n!\,(n-i)}{(n-i)!\,i!\,(i+1)} p^i (1-p)^{n-i} \frac{p}{1-p}$$

$$= \frac{n-i}{i+1} \frac{p}{1-p} p^i$$

Variabel Acak Poisson

Sebuah variabel acak X yang menggunakan salah satu nilai 0, 1, 2 ... dikatakan merupakan variabel acak Poisson dengan parameter λ , $\lambda > 0$, jika fungsi massa probabilitasnya ditentukan oleh

$$p_i = P\{X = i\} = e^{-\lambda} \frac{\lambda^i}{i!}, \qquad i = 0, 1, \ldots$$

Simbol e, yang didefinisikan dengan $e = \lim_{n \to \infty} (1 + 1/n)^n$, merupakan konstan terkenal dalam matematika yang kira-kira sama dengan 2,7183.

Variabel acak Poisson mempunyai jajaran luas aplikasi. Salah satu alasan untuk ini adalah bahwa variabel acak seperti ini dapat digunakan untuk mengaproksimasi distribusi jumlah keberhasilan dalam jumlah besar percobaan (yang mungkin independen atau paling baik "agak dependen") ketika tiap percobaan mempunyai probabilitas kecil untuk berhasil. Untuk melihat mengapa demikian halnya, andaikan saja bahwa X adalah variabel acak binomial dengan parameter (n, p) — dan dengan demikian merepresentasikan jumlah keberhasilan dalam n

percobaan independen ketika masing-masing percobaan berhasil dengan probabilitas p—dan misalkan $\lambda = np$. Maka

$$
\begin{aligned}
P\{X=i\} &= \frac{n!}{(n-i)!\,i!}\, p^{i}(1-p)^{n-1} \\
&= \frac{n!}{(n-i)!\,i!}\left(\frac{\lambda}{n}\right)^{i}\left(1-\frac{\lambda}{n}\right)^{n-1} \\
&= \frac{n(n-1)\ldots(n-i+1)}{n^{i}}\;\frac{\lambda^{i}}{i!}\;\frac{(1-\lambda/n)^{n}}{(1-\lambda/n)^{i}}
\end{aligned}
$$

Sekarang untuk n besar dan p kecil,

$$
\left(1-\frac{\lambda}{n}\right)^{n} \approx e^{-\lambda}, \quad \frac{n(n-1),\ldots(n-i+1)}{n^{i}} \approx 1, \quad \left(1-\frac{\lambda}{n}\right)^{i} \approx 1
$$

Oleh karena itu, untuk n besar dan p kecil,

$$
P\{X=i\} \approx e^{-\lambda}\,\frac{\lambda^{i}}{i!}
$$

Karena rerata dan variansi variabel acak binomial Y ditentukan oleh

$$
E[Y] = np, \quad \mathrm{Var}(Y) = np(1-p) \approx np \qquad \text{untuk } p \text{ kecil}
$$

maka adalah berdasarkan intuisi, dengan hubungan antara variabel acak binomial dan Poisson, bahwa untuk variabel acak poisson X, yang mempunyai parameter λ,

$$
E[X] = \mathrm{Var}(X) = \lambda
$$

Suatu bukti analitis dari persamaan di atas disisakan sebagai latihan.

Untuk menghitung probabilitas Poisson kita menggunakan rumus rekursif berikut ini:

$$
\frac{p_{i+1}}{p_i} = \frac{\dfrac{e^{-\lambda}\lambda^{i+1}}{(i+1)!}}{\dfrac{e^{-\lambda}\lambda^{i}}{i!}} = \frac{\lambda}{i+1}
$$

atau, sama dengan,

$$p_{i+1} = \frac{\lambda}{i+1} p^{i}, \qquad i \geq 0$$

Variabel Acak Geometrik

Perhatikan percobaan independen, yang masing-masing berhasil dengan probabilitas p. Jika X mewakili bilangan percobaan pertama yang berhasil maka

$$P\{X = n\} = p(1 - p)^{n-1}, n \geq 1 \tag{2.7}$$

yang dapat dengan mudah diperoleh dengan memperhatikan bahwa agar keberhasilan pertama terjadi pada percobaan ke-n maka n -1 yang pertama semuanya harus lebih dahulu berupa kegagalan dan yang ke-n keberhasilan. Persamaan (2.7) sekarang menyusul karena percobaan-percobaan tersebut independen.

Variabel acak yang fungsi massa probabilitasnya ditentukan oleh (2.7) dikatakan merupakan variabel acak geometrik dengan parameter p. Rerata dari geometrik tersebut diperoleh sebagai berikut:

$$E[X] \sum_{n=1}^{\infty} np(1-p)^{n-1} = \frac{1}{p}$$

di mana persamaan di atas menggunakan identitas aljabar, untuk $0 < x < 1$,

$$\sum_{n=1}^{\infty} nx^{n-1} = \frac{d}{dx}\left(\sum_{n=0}^{\infty} x^{n}\right) = \frac{d}{dx}\left(\frac{1}{1-x}\right) = \frac{1}{(1-x)^2}$$

Juga tidak sulit untuk memperlihatkan bahwa

$$\mathrm{Var}(X) = \frac{1-p}{p^2}$$

Variabel Acak Binomial Negatif

Jika kita misalkan X menandai jumlah percobaan yang diperlukan untuk mengumpulkan total keberhasilan r ketika tiap percobaan secara independen merupakan keberhasilan dengan probabilitas p, maka X dikatakan sebagai binomial negatif, kadang disebut Pascal, variabel acak dengan parameter p dan r. Fungsi massa probabilitas variabel acak seperti ini diberikan sebagai berikut:

$$P\{X=n\}=\binom{n-1}{r-1}p^r(1-p)^{n-r}, \qquad n \geq r \tag{2.8}$$

Untuk melihat mengapa Persamaan (2.8) absah, perhatikan bahwa agar persamaan itu melakukan n percobaan untuk menghimpun s keberhasilan, n pertama - 1 percobaan harus menghasilkan persis r - 1 keberhasilan — dan probabilitasnya adalah $\binom{n-1}{r-1}p^{r-1}(1-p)^{n-r}$ — dan kemudian percobaan ke-n pasti berhasil — dan probabilitasnya adalah p.

Jika X_i, $i = 1, \ldots, r$, menandakan jumlah percobaan yang diperlukan sesudah keberhasilan ke-(i - 1) untuk mendapatkan keberhasilan ke-i, maka mudahlah untuk melihat bahwa mereka merupakan variabel acak geometrik yang independen dengan parameter bersama p. Karena

$$X=\sum_{i=1}^{r} X_i$$

kita lihat bahwa

$$E[X]=\sum_{i=1}^{r} E[X_i]=\frac{r}{p}$$

$$Var(X)=\sum_{i=1}^{r} Var(X_i)=\frac{r(1-p)}{p^2}$$

di mana rumus di atas menggunakan hasil yang sesuai untuk variabel acak geometrik.

Variabel Acak Hipergeometrik

Perhatikan sebuah ember yang berisi $N + M$ bola. N berwarna terang dan M berwarna gelap. Jika sampel ukuran n dipilih secara acak [dalam pengertian bahwa tiap himpunan bagian $\binom{N+M}{n}$ ukuran n sama mungkinnya untuk dipilih] maka X, jumlah bola berwarna terang yang dipilih, memiliki fungsi massa probabilitas

$$P\{X=i\}=\frac{\binom{N}{i}\binom{M}{n-i}}{\binom{N+M}{n}}$$

Variabel acak X yang fungsi massa probabilitasnya ditentukan oleh persamaan di atas disebut variabel acak hipergeometrik.

Andaikan n bola dipilih secara berurutan. Jika

$$X_i = \begin{cases} 1 & \text{jika seleksi ke } i \text{ adalah terang} \\ 0 & \text{lainnya} \end{cases}$$

maka

$$X = \sum_{i=1}^{n} X_i \quad (2.9)$$

dan dengan demikian

$$E[X] = \sum_{i=1}^{n} E[X_i] = \frac{nN}{N+M}$$

di mana persamaan di atas menggunakan fakta bahwa, dengan simetri, seleksi ke-i, seleksi ke-i sama mungkinnya berupa $N + M$ bola, dan dengan demikian $E[X_i] = P\{X_i = 1\} = N/(N + M)$.

Karena X_i tidak independen (mengapa tidak?), pemanfaatan representasi (2.9) untuk menghitung Var(X) melibatkan istilah kovarian. Produk akhirnya dapat diperlihatkan untuk memberikan hasil

$$\text{Var}(X) = \frac{nNM}{(N+M)^2}\left(1 - \frac{n-1}{N+M-1}\right)$$

2.9. Variabel Acak Kontinu

Dalam bagian ini kita mempertimbangkan jenis-jenis tertentu dari variabel acak kontinu.

Variabel Acak Terdistribusi Seragam

Variabel acak X dikatakan terdistribusi seragam atas interval (a, b), $a < b$, jika fungsi densitas probabilitasnya diberikan oleh

$$f(x) = \begin{cases} \dfrac{1}{(b-a)} & \text{jika } a < x < b \\ 0 & \text{lainnya} \end{cases}$$

Dengan kata lain, X terdistribusi seragam atas (a, b) jika menempatkan semua massanya pada interval itu dan sama mungkinnya berada "dekat" titik mana pun pada interval itu.

Rerata dan variansi suatu variabel acak (a, b) seragam diperoleh sebagai berikut:

$$E[X] = \frac{1}{b-a}\int_a^b x\,dx = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2}$$

$$E[X^2] = \frac{1}{b-a}\int_a^b x^2\,dx = \frac{b^3-a^3}{3(b-a)} = \frac{a^2+b^2+ab}{3}$$

dan dengan begitu

$$\text{Var}(X) = \tfrac{1}{3}(a^2+b^2+ab) - \tfrac{1}{4}(a^2+b^2+2ab) = \tfrac{1}{12}(b-a)^2$$

Jadi sebagai contoh, nilai yang diharapkan adalah, seperti yang mungkin diduga, titik tengah interval (a, b).

Fungsi distribusi X diberikan, untuk $a < x$, b, dengan

$$F(x) = P\{X \le x\} = \int_a^x (b-a)^{-1}\,dx = \frac{x-a}{b-a}$$

Variabel Acak Normal

Variabel acak X dikatakan terdistribusi normal dengan rerata μ dan variansi σ^2 jika fungsi densitas probabilitasnya diberikan oleh

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2}, \qquad -\infty < x < \infty$$

Densitas normal merupakan kurva bentuk lonceng yang simetris sekitar μ (lihat Gambar 2.1).

Tidak sulit untuk memperlihatkan bahwa parameter μ dan σ^2 benar-benar sama dengan harapan dan variansi yang normal. Yaitu,

$$E[X] = \mu \qquad \text{dan} \qquad Var(X) = \sigma^2$$

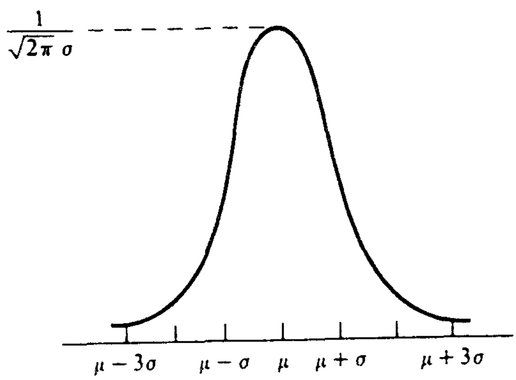
Fakta penting tentang variabel acak normal adalah bahwa jika X normal dengan rerata μ dan variansi σ^2, maka untuk tiap konstan a dan b, $aX + b$ terdistribusi secara normal dengan rerata $a\mu + b$ dan variansi $a^2\,\sigma^2$ Jika X normal dengan rerata μ dan variansi σ maka

$$Z = \frac{X - \mu}{\sigma}$$

normal dengan rerata 0 dan variansi 1. Variabel acak Z seperti ini dikatakan memiliki distribusi normal per unit. Misalkan ϕ menandakan fungsi distribusi variabel acak normal per unit; yaitu,

$$\phi(x) = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{x} e^{-x^2/2}\, dx, \quad -\infty < x < \infty$$

Hasil bahwa $Z = (X - \mu)/\sigma$ memiliki distribusi normal per unit ketika X normal dengan rerata μ dan variansi σ^2 sangat berguna karena memungkinkan kita mengevaluasi semua probabilitas sehubungan dengan X dalam kaitan dengan ϕ Sebagai contoh, fungsi distribusi X dapat dinyatakan sebagai



Gambar 2.1 Fungsi densitas normal.

$$
\begin{aligned}
F(x) &= P\{X \le x\} \\
&= P\left\{\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right\} \\
&= P\left\{Z \le \frac{x-\mu}{\sigma}\right\} \\
&= \phi\left(\frac{x-\mu}{\sigma}\right)
\end{aligned}
$$

Nilai $\phi(x)$ dapat ditentukan dengan melihatnya dalam tabel atau dengan menuliskan sebuah program komputer untuk memperkirakannya.

Untuk a dalam interval (0, 1), maka z_a menjadi

$$P\{Z > z_a\} = 1 - \phi(z_a) = a$$

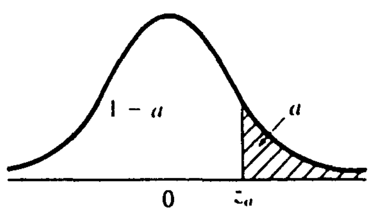
Dengan kata lain, unit normal akan melebihi z_a dengan probabilitas a. (lihat Gambar 2.2) Nilai z_a dapat diperoleh dari tabel nilai ϕ. Sebagai contoh, karena

$$\phi(1.64) = .95, \qquad \phi(1.96) = .975, \qquad \phi(2.33) = .99$$

kita lihat bahwa

$$z_{.05} = 1.64, \qquad z_{.025} = 1.96, \qquad z_{.01} = 2.33$$

Aplikasi luas variabel acak normal dihasilkan dari salah satu dalil terpenting dalam teori probabilitas — dalil limit pusat, yang menegaskan bahwa jumlah dari



Gambar 2.2 $P\{Z > z_a\} = a$

banyak sekali variabel acak independen memiliki distribusi yang mendekati normal. Bentuk paling sederhana dari dalil yang luar biasa ini adalah sebagai berikut.

Dalil Limit Pusat

Misalkan $X_1, X_2, \ldots$ merupakan urutan variabel acak terdistribusi identik yang memiliki rerata .. terbatas dan variansi .. terbatas. Maka

$$\lim_{n\to\infty} P\left\{ \frac{X_1 + \ldots + X_n - n\mu}{\sigma\sqrt{n}} < x \right\} = \phi(x)$$

Variabel Acak Eksponensial

Variabel acak kontinu yang memiliki fungsi densitas probabilitas

$$f(x) = \lambda e^{-\lambda x}, \qquad 0 < x < \infty$$

untuk $\lambda > 0$ dikatakan merupakan variabel acak eksponensial dengan parameter λ. Distribusi kumulatifnya diberikan dengan

$$F(x) = \int_0^x \lambda e^{-\lambda x}\, dx = 1 - e^{-\lambda x}, \; 0 < x < \infty$$

Mudah untuk memverifikasi bahwa nilai yang diharapkan dan variansi variabel acak seperti ini adalah sebagai berikut:

$$E[X] = \frac{1}{\lambda} \quad \text{dan} \quad \mathrm{Var}(X) = \frac{1}{\lambda^2}$$

Sifat pokok variabel acak eksponensial adalah variabel ini memiliki "sifat tanpa memori." Kita mengatakan bahwa variabel acak nonnegatif adalah tanpa memori jika

$$P\{X > s + t \mid X > s\} = P\{X > t\} \qquad \text{untuk semua } s, \tau \geq 0 \qquad (2.10)$$

Untuk mengerti mengapa persamaan di atas disebut sifat tanpa memori, bayangkan bahwa X mewakili jangka hidup suatu unit, dan perhatikan probabilitas bahwa suatu unit berusia s akan bertahan selama waktu tambahan t. Karena hal ini akan terjadi jika jangka hidup unit bersangkutan melebihi $t + s$ seandainya masih hidup pada waktu s, kita lihat bahwa

$$P\{\text{tambahan hidup suatu item usia } s \text{ melebihi } t\} = P\{X > s + t \mid X > s\}$$

Persamaan (2.10) merupakan pernyataan fakta bahwa distribusi hidup selebihnya dari suatu item usia s tidak bergantung pada s. Dengan kata lain, tidak perlu untuk mengingat usia unit bersangkutan untuk mengetahui distribusinya atas hidup yang selebihnya.

Persamaan: (2.10) sama dengan

$$P\{X > s + t\} = P\{X > s\}\ P\{X > t\}$$

Persamaan di atas dipenuhi setiap kali X merupakan variabel acak eksponensial — karena, dalam hal ini, $P\ \{X > x\} = e^{-\lambda x}$ — kita lihat bahwa variabel acak eksponensial adalah tanpa memori (dan sebenarnya tidak sulit untuk memperlihatkan bahwa variabel tersebut merupakan satu-satunya variabel acak yang tanpa memori).

Satu sifat lain yang berguna dari variabel acak eksponensial adalah variabel ini tetap merupakan eksponensial ketika dikali dengan konstan positif. Untuk melihat ini, andaikan saja bahwa adalah eksponensial dengan parameter λ, dan c merupakan bilangan positif. Maka

$$P\{cX \leq x\} = P\left\{X \leq \frac{x}{c}\right\} = 1 - e^{-\lambda x/c}$$

yang memperlihatkan bahwa cX adalah eksponensial dengan parameter λ/c.

Proses Poisson dan Variabel Acak Gamma

Andaikan bahwa "peristiwa" terjadi pada titik waktu acak dan $N(t)$ menandakan jumlah peristiwa yang terjadi pada interval waktu $[0, t]$. Peristiwa-peristiwa ini dikatakan merupakan *proses Poisson yang memiliki nilai* λ, $\lambda > 0$, jika

(a) $N(0) = 0$

(b) Jumlah peristiwa yang terjadi dalam interval waktu terputus adalah independen.

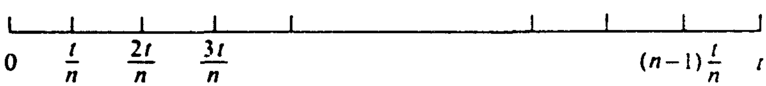
(c) Distribusi jumlah peristiwa yang terjadi dalam interval tertentu bergantung hanya pada panjangnya interval dan bukan pada lokasinya.

(d) $\lim_{h \to 0} \dfrac{P\{N(h) = 1\}}{h} = \lambda$

(e) $\lim_{h \to 0} \dfrac{P\{N(h) \geq 2\}}{h} = 0$

Jadi, Kondisi (a) menyatakan bahwa prosesnya dimulai pada waktu 0. Kondisi (b), asumsi *inkremen independen*, menyatakan bahwa jumlah peristiwa pada waktu t\{yaitu, $N(t)$] independen terhadap jumlah peristiwa yang terjadi antara t dan $t + s$ [yaitu, $N(t + s) - N(t)$]. Kondisi (c), asumsi *inkremen stasioner*, menyatakan bahwa distribusi probabilitas $N(t + s) - N(t)$ sama dengan semua nilai t. Kondisi (d) dan (e) menyatakan bahwa dalam interval kecil panjang h, probabilitas terjadinya satu peristiwa kira-kira λh, sementara probabilitas terjadinya dua atau lebih peristiwa kira-kira 0.

Kita sekarang berargumen bahwa asumsi ini menyiratkan bahwa jumlah peristiwa yang terjadi dalam interval panjang t adalah variabel acak Poisson dengan rerata λt. Untuk melakukan ini, perhatikan interval $[0, t]$, dan pecahlah menjadi menjadi n subinterval yang tidak tumpang tindih dari panjang t/n (Gambar 2.3). Perhatikan lebih dahulu jumlah subinterval ini yang mengandung suatu peristiwa. Karena setiap subinterval secara independen [dengan Kondisi (b)] mengandung suatu peristiwa dengan probabilitas yang sama [dengan Kondisi (c)], yang kira-kira sama dengan $\lambda t/n$, maka jumlah interval seperti ini adalah variabel acak binomial dengan parameter n dan $p \approx \lambda t/n$. Oleh karena itu, dengan argumen yang menghasilkan penyatuan binomial dengan Poisson, kita melihat dengan $n \to \infty$ bahwa jumlah subinterval seperti ini menyatu dengan variabel acak Poisson dengan rerata λt. Seperti terlihat Kondisi (e) menyiratkan probabilitas bahwa yang mana pun dari subinterval ini mengandung dua atau lebih peristiwa menuju



Gambar 2.3

0 sebagai $n \to \infty$, maka $N(t)$, jumlah peristiwa yang terjadi dalam $[0, t]$, merupakan variabel acak Poisson dengan rerata λt.

Untuk proses Poisson, mari kita tandai waktu peristiwa pertama dengan X_1. Lebih jauh, untuk $n > 1$, X_n menandakan waktu yang berlalu antara peristiwa ke-$(n - 1)$ dan peristiwa ke-n. Urutan $\{X_n, n = 1, 2, \ldots\}$ disebut *urutan waktu antarkedatangan*. Sebagai contoh, jika $X_1 = 5$ dan $X_2 = 10$, maka peristiwa pertama dalam proses Poisson akan terjadi pada waktu 5 dan kedua pada waktu 15.

Kita sekarang menentukan distribusi X_n. Untuk melakukan itu, kita lebih dahulu memperhatikan bahwa peristiwa $\{X_1 > t\}$ terjadi jika dan hanya jika tidak ada peristiwa dalam proses Poisson terjadi pada interval $[0, t]$, sehingga

$$P\{X_1 > t\} = P\{N(t) = 0\} = e^{-\lambda t}$$

Oleh karena itu X_1 memiliki distribusi eksponensial dengan rerata $1/\lambda$. Untuk mendapatkan distribusi X_2, perhatikan bahwa

$$\begin{aligned} P\{X_2 > t | X_1 = s\} &= P\{0 \text{ events in } (s, \; s + t] | X_1 = s\} \\ &= P\{0 \text{ events in } (s, \; s + t]\} \\ &= e^{-\lambda t} \end{aligned}$$

di mana dua persamaan terakhir menyusul dari inkremen independen dan stasioner. Oleh karena itu, dari yang sebelumnya ini, kita menyimpulkan bahwa X_2 juga merupakan variabel acak eksponensial dengan rerata $1/\lambda$. dan, lebih jauh, bahwa X_2 independen X_1. Pengulangan argumen yang sama menghasilkan:

Proposisi

Waktu antarinterval $X_1, X_2, \ldots$ merupakan variabel acak eksponensial yang independen dan terdistribusi secara identik dengan parameter λ.

Misal $S_n = \sum_{i=1}^{n} X_i$ menandakan waktu peristiwa ke-n. Karena S_n akan kurang dari atau sama dengan t jika dan hanya jika sedikitnya ada n peristiwa pada waktu t, kita lihat bahwa

$$P\{S_n \le t\} = P\{N(t) \ge n\}$$

$$= \sum_{j=n}^{\infty} e^{-\lambda t} \frac{(\lambda t)^j}{j!}$$

Karena sisi kiri merupakan fungsi distribusi kumulatif dari S_n, kita mendapatkan, dengan diferensiasi, bahwa fungsi densitas S_n — menyebutnya $f_n(t)$ — diberikan dengan

$$\begin{aligned} f_n(t) &= \sum_{j=n}^{\infty} \lambda e^{-\lambda t} \frac{j(\lambda t)^{j-1}}{j!} - \sum_{j=n}^{\infty} \lambda e^{-\lambda t} \frac{(\lambda t)^j}{j!} \\ &= \sum_{j=n}^{\infty} \lambda e^{-\lambda t} \frac{(\lambda t)^{j-1}}{(j-1)!} - \sum_{j=n}^{\infty} \lambda e^{-\lambda t} \frac{(\lambda t)^j}{j!} \\ &= \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} \end{aligned}$$

Definisi

Variabel acak yang memiliki fungsi densitas probabilitas

$$f(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}, \qquad t > 0$$

dikatakan sebagai variabel acak gamma dengan parameter (n, λ).

Jadi kita lihat bahwa S_n, waktu dari peristiwa ke-n dalam proses Poisson yang memiliki nilai λ, merupakan variabel acak gamma dengan parameter (n, λ). Sebagai tambahan, kita mendapatkan dari representasi $S_n = \sum_{i=1}^{n} X_i$ dan proposisi sebelumnya, yang menyatakan bahwa X_i ini merupakan eksponensial independen dengan nilai λ, konsekuensi wajar berikut ini.

Konsekuensi wajar:

Jumlah n variabel acak eksponensial independen, yang masing-masing memiliki paramter λ, adalah variabel acak gamma dengan parameter (n, λ).

Proses Poisson Nonhomogen

Dari sudut pandang pemodelan, kelemahan utama dari proses Poisson adalah asumsinya bahwa peristiwa benar-benar mungkin terjadi pada semua interval dalam ukuran yang sama. Generalisasi, yang melonggarkan asumsi ini, menghasilkan proses yang disebut nonhomogen atau nonstasioner.

Jika "peristiwa" terjadi pada waktu acak, dan $N(t)$ menandakan jumlah peristiwa yang terjadi pada waktu t, maka kita katakan bahwa $\{N(t)\ t \geq 0\}$ merupakan proses Poisson nonhomogen dengan fungsi intensitas $\lambda(t)$, $t \geq 0$, jika

(a) $N(0) = 0$.
(b) Jumlah peristiwa yang terjadi dalam interval waktu terputus adalah independen.
(c) $\lim_{h \to 0} P$ {persis 1 peristiwa antara t dan $t + h\}/h = \lambda(t)$
(d) $\lim_{h \to 0} P\{2$ atau lebih peristiwa antara t dan $+ h\}/h = 0$

Fungsi $m(t)$, yang didefinisikan dengan

$$m(t) = \int_0^t \lambda(s)\, ds, \qquad t \geq 0$$

disebut fungsi nilai-rerata. Hasil berikut ini dapat ditetapkan.

Proposisi:

$N(t + s) - N(t)$ adalah variabel acak Poisson dengan rerata $m(t + s) - m(t)$.

Kuantitas $\lambda(t)$, yang disebut intensitas pada waktu t, menunjukkan seberapa mungkin bahwa suatu peristiwa akan terjadi sekitar waktu t. [Perhatikan bahwa ketika $\lambda(t) \equiv. \lambda$ proses yang nonhomogen kembali pada proses Poisson yang biasa.] Proposisi berikut ini memberikan cara yang berguna untuk menafsirkan proses Poisson yang nonhomogen.

Proposisi:

Andaikan bahwa peristiwa terjadi menurut proses Poisson yang memiliki nilai λ, dan andaikan bahwa, lepas dari apa pun yang datang sebelumnya, suatu peristiwa yang terjadi pada waktu t dihitung dengan probabilitas $p(t)$. Maka, proses peristiwa terhitung merupakan proses Poisson nonhomogen dengan fungsi intensitas $\lambda(t) = \lambda p(t)$.

Bukti. Proposisi ini dibuktikan dengan memperhatikan bahwa kondisi tertentu sebelumnya semua dipenuhi. Kondisi (a), (b), dan (d) menyusul karena hasil yang sesuai berlaku untuk semua peristiwa (bukan hanya yang terhitung). Kondisi (c) menyusul karena

$$\begin{aligned} P\{1 \text{ peristiwa terhitung antara } t \text{ dan } t+h\} \\ &= P\{1 \text{ peristiwa dan terhitung}\} \\ &+ P\{2 \text{ atau lebih peristiwa dan hanya 1 terhitung}\} \\ &\approx \lambda h p(t) \end{aligned}$$

2.10. Harapan Bersyarat dan Variansi Bersyarat

Jika X dan Y adalah variabel acak diskret bersama, kita definisikan $E[X| Y = y]$, harapan bersyarat X seandainya $Y = y$, dengan

$$\begin{aligned} E[X \mid Y = y] &= \sum_x xP\{X = x \mid Y = y\} \\ &= \frac{\sum_x xP\{X = x, Y = y\}}{P\{Y = y\}} \end{aligned}$$

Dengan kata lain, harapan bersyarat X, dengan $Y = y$, didefinisikan seperti sebelumnya sebagai rata-rata tertimbang dari semua nilai yang mungkin dari X tetapi sekarang dengan bobot yang diberikan pada nilai x yang sama dengan probabilitas bersyarat bahwa X sama dengan x seandainya Y sama dengan y.

Demikian pula halnya, jika X dan Y berkesinambungan bersama dengan fungsi densitas bersama $f(x, y)$, kita definisikan harapan bersyarat X, seandainya $Y = y$, dengan

$$E[X \mid Y = y] = \frac{\int x f(x, y)\, dx}{\int f(x, y)\, dx}$$

Misalkan $E[X \mid Y]$ menandakan bahwa fungsi variabel acak Y yang nilainya pada $Y = y$ adalah $E[X \mid Y = y]$; dan perhatikan bahwa $E[X \mid Y]$ itu sendiri adalah sebuah variabel acak. Proposisi berikut ini sangat berguna.

Proposisi:

$$E[E[X \mid Y]] = E[X] \tag{2.11}$$

Jika y merupakan variabel acak diskret maka Persamaan (2.11) menyatakan bahwa

$$E[X] = \sum_y E[X \mid Y = y] P\{Y = y\}$$

sementara jika Y berkesinambungan dengan densitas g, maka (2.11) menyatakan

$$E[X] = \int E[X \mid Y = y] g(y)\, dy$$

Kita sekarang memberikan bukti untuk proposisi di atas jika X dan Y bersifat diskret:

$$\begin{aligned}
\sum_y E[X \mid Y = y] P\{y = y\} &= \sum_y \sum_x xP\{X = x \mid Y = y\} P\{Y = y\} \\
&= \sum_y \sum_x xP\{X = x, Y = y\} \\
&= \sum_y x \sum_x P\{X = x, Y = y\} \\
&= \sum_y xP\{X = x\} \\
&= E[X]
\end{aligned}$$

Kita juga dapat mendefinisikan variansi bersyarat X, dengan nilai Y, sebagai berikut:

$$\text{Var}(X \mid Y) = E[(X - E\{X \mid Y])^2 \mid Y]$$

Dengan kata lain, $\text{Var}(X \mid Y)$ adalah fungsi Y, yang pada $Y = y$ sama dengan variansi X seandainya $Y = y$. Dengan alasan yang sama yang menghasilkan identitas $\text{Var}(X) = E[X^2] - (E[X])^2$ kita mendapatkan

$$\text{Var}(X \mid Y) = E[X^2 \mid Y] - (E[X \mid Y])^2$$

Dengan harapan dari kedua sisi persamaan di atas ditetapkan

$$\begin{aligned} E[\text{Var}(X \mid Y)] &= E[E[X^2 \mid Y]] - E[(E[X \mid Y])^2] \\ &= E[X^2] - E[(E[X \mid Y])^2] \end{aligned} \tag{2.12}$$

Demikian juga, karena $E[E[X \mid Y]] = E[X]$, kita mendapatkan

$$\text{Var}(E[X \mid Y]) = E[(E[X \mid Y])^2] - (E[X])^2 \tag{2.13}$$

Dengan menambahkan Persamaan (2.12) dan (2.13) kita mendapatkan identitas berikut ini, yang dikenal sebagai rumus variansi bersyarat.

RUMUS VARIANSI BERSYARAT

$$\text{Var}(X) = E[\text{Var}(X \mid Y)] + \text{Var}(E[X \mid Y])$$

Soal

1. Ada sebuah eksperimen dengan enam ekor kuda, yang diberi nomor 1 hingga 6. Kuda-kuda ini diperlombakan, dan andaikan saja ruang sampel ditetapkan dengan

   $$S = \{\text{semua urutan } (1, 2, 3, 4, 5, 6)\}$$

   Misalkan A menandakan peristiwa di mana kuda nomor 1 berada di antara tiga teratas yang tiba di garis finis, misalkan B menandakan peristiwa di mana kuda nomor 2 tiba di tempat kedua, dan misalkan C menandakan peristiwa di mana kuda nomor 3 tiba di tempat ketiga.

   (a) Uraikan peristiwa $A \cup B$. Berapa banyak hasil terkandung di dalam peristiwa ini?

   (b) Berapa banyak hasil terkandung dalam peristiwa AB?

(c) Berapa banyak hasil terkandung dalam peristiwa ABC?

(d) Berapa banyak hasil terkandung dalam peristiwa $A \cup BC$?

2. (a) Untuk peristiwa A dan B mana pun perlihatkan bahwa

$$A \cup B = A \cup A^c B$$
$$A = AB \cup AB^c$$

(b) Perlihatkan bahwa $P(A \cup B) = P(A) + P(B) - P(AB)$

3. Suatu pasangan mempunyai dua orang anak. Bagaimana dengan probabilitas bahwa keduanya anak perempuan seandainya anak yang tertua adalah anak perempuan? Asumsikanlah bahwa keempat kemungkinan sama besarnya.

4. Raja berasal dari sebuah keluarga dengan dua anak. Bagaimana probabilitas bahwa anak yang satunya adalah saudaranya?

5. Variabel acak X menggunakan salah satu dari nilai 1, 2, 3, 4 dengan probabilitas

$$P\{X = i\} = ic, \qquad i = 1,2,3,4$$

untuk nilai c. Temukan $P\{2 \leq X \leq 3\}$

6. Variabel acak kontinu X memiliki fungsi densitas probabiliats yang ditetapkan dengan

$$f(x) = cx, \qquad 0 < x < 1$$

Temukan $P\{X > \frac{1}{2}\}$

7. Jika X dan Y memiliki fungsi densitas probabilitas bersama yang ditetapkan dengan

$$f(x,y) = 2e\ {-(x + 2y)}, \qquad 0 < x < \infty\ , 0 < y < \infty$$

Temukan $P\{X < Y\}$.

8. Temukan nilai yang diharapkan dari variabel acak yang ditetapkan dalam Soal 5.

9. Temukan $E[X]$ untuk variabel acak Soal 6.

10. Ada 10 jenis kupon yang berbeda dan tiap kali seseorang mendapatkan selembar kupon maka ada kemungkinan sama besar bahwa kupon itu adalah salah satu dari ke-10 jenis. Misalkan X menandakan jumlah jenis yang berbeda yang terkandung di dalam sekumpulan N kupon, dan temukan $E[X]$. [Petunjuk: Untuk $i = 1, \ldots, 10$ misalkan

$$X_i = \begin{cases} 1 & \text{jika kupon sejenis } i \text{ ada diantara} N \\ 0 & \text{lainnya} \end{cases}$$

dan gunakan representasi $X = \sum_{i=1}^{10} x_i$.

11. Sebuah dadu yang mempunyai enam muka digulirkan. Jika masing-masing dari enam hasil yang mungkin keluar sama mungkinnya, tentukan variansi angka yang muncul.

12. Andaikan bahwa X memiliki fungsi densitas probabilitas

$$f(x) = ce^x, \qquad 0 < x < 1$$

Tentukan Var(X).

13. Perlihatkan bahwa Var($aX + b$) = a^2Var(X).

14. Andaikan bahwa X, jumlah apel cair yang ada di dalam wadah sari buah apel, adalah variabel acak yang memiliki rerata 4 gram.

(a) Apa pendapat anda tentang probabilitas bahwa wadah tertentu berisikan lebih dari 6 gram sari buah apel?

(b) Jika (Var(X) = 4(gram)^2, apa yang dapat dikatakan tentang probabilitas bahwa wadah tertentu akan berisikan antara 3 dan 5 gram sari buah apel?

15. Sebuah pesawat terbang memerlukan sedikitnya setengah dari mesin-mesinnya untuk menyelesaikan perjalanannya dengan selamat. Jika tiap mesin berfungsi sendiri-sendiri dengan probabilitas p, dengan nilai p berapakah sebuah pesawat terbang bermesin tiga lebih aman dibandingkan pesawat bermesin lima?

16. Untuk variabel acak binomial X dengan parameter (n, p), perlihatkan bahwa $P\{X = i\}$ lebih dahulu bertambah dan kemudian berkurang, mencapai nilai maksimumnya ketika i adalah bilangan bulat terbesar yang kurang dari atau sama dengan $(n + 1)p$.

17. Jika X dan Y adalah variabel acak binomial independen dengan parameter masing-masing (n, p) dan (m, p), buktikan tanpa perhitungan apa pun, bahwa $X + Y$ adalah binomial dengan parameter $(n + m, p)$.

18. Jelaskan mengapa variabel acak berikut ini semuanya mempunyai distribusi Poisson:

(a) Jumlah kesalahan cetak pada bab tertentu buku ini.
(b) Jumlah nomor telepon yang keliru diputar setiap hari.
(c) Jumlah pelanggan yang memasuki kantor pos tertentu pada hari tertentu.

19. Jika X adalah variabel acak Poisson dengan parameter λ, perlihatkan bahwa
(a) $E[X] = \lambda$
(b) $\mathrm{Var}(X) = \lambda$

20. Misalkan X dan Y merupakan variabel acak Poisson yang independen dengan parameter masing-masing λ_1 dan λ_2. Gunakan hasil Soal 17 untuk membuktikan dengan penyelidikan sendiri bahwa $X + Y$ adalah Poisson dengan parameter $\lambda_1 + \lambda_2$. Lalu berikanlah sebuah bukti analitis tentang hal ini. [Petunjuk:

$$P\{X+Y=k\} = \sum_{i=0}^{k} P\{X=i, Y=k-i\} = \sum_{i=0}^{k} P\{X=i\}P\{Y=k-i\}]$$

21. Jelaskan bagaimana memanfaatkan hubungan

$$p(i+1) = \frac{\lambda}{i+1}p(i)$$

untuk menghitung secara efisien probabilitas Poisson.

22. Temukan $P\{X > n\}$ jika X adalah sebuah variabel acak eometrik dengan parameter p.

23. Dua orang pemain memainkan permainan tertentu hingga salah satu memenangkan total lima permainan. Jika pemain A memenangkan tiap permainan indvidual dengan probabilitas 0,6, maka berapakah probabilitas ia akan memenangkan pertandingan tersebut?

24. Perhatikan model hipergeometrik Bagian 2.8, dan andaikan bahwa semua bola putihnya diberi nomor. Untuk $i = 1, \ldots, N$ misalkan

$$Y_i = \begin{cases} 1 & \text{jika bola putih bernomor } i \text{ dipilih} \\ 0 & \text{lainnya} \end{cases}$$

Buktikan bahwa $X = \sum_{i=1}^{N} Y_i$ dan kemudian gunakan representasi ini untuk menentukan $E[X]$. Buktikan bahwa ini cocok dengan hasil yang diberikan dalam Bagian 2.8.

25. Bus akan tiba pada waktu yang didistribusikan secara seragam antara pukul 8.00 dan 8.30. Jika kita tiba pada pukul 8.00, berapa besar probabilitas bahwa kita akan menunggu antara 5 dan 15 menit?

26. Untuk variabel acak normal dengan parameter λ dan σ^2 perlihatkan bahwa

(a) $E[X] = \lambda$
(b) $\text{Var}(X) = \sigma^2$

27. Misalkan X merupakan variabel acak binomial dengan parameter (n, p). Jelaskan mengapa

$$P\left\{\frac{X - np}{\sqrt{np(1 - p)}} \leq x\right\} \approx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-x^2/2}\, dx$$

ketika n besar.

28. Jika X merupakan variabel acak eksponensial dengan parameter λ, perlihatkan bahwa
(a) $E[X] = 1/\lambda$.
(b) $\text{Var}(X) = 1/\lambda^2$.

29. Orang A, B, dan C menunggu di bank yang dilayani dengan dua orang kasir ketika buka pada pagi hari. Orang A dan B masing-masing menghampiri seorang kasir dan C menunggu di barisan. Jika waktu yang dibutuhkan untuk melayani seorang pelanggan merupakan variabel acak eksponensial dengan parameter λ, berápa besar probabilitas C sebagai orang terakhir yang meninggalkan bank? [Petunjuk: Tidak perlu penghitungan sama sekali.]

30. Pertimbangkan proses Poisson di mana peristiwa terjadi pada kecepatan 0,3 per jam. Bagaimana probabilitas tidak adanya peristiwa yang terjadi antara pukul 10.00 dan pukul 14.00?

31. Untuk proses Poisson dengan kecepatan λ, temukan $P\{N(s) = k | N(t) = n\}$ jika $s < t$.

32. Ulangi Soal 31 untuk $s > t$.

33. Jika X adalah variabel acak gamma dengan paramter (n, p), dapatkan
(a) $E[X]$.
(b) $\text{Var}(X)$.

34. Sebuah bejana berisi empat buah bola putih dan enam buah bola hitam. Sebuah sampel acak ukuran 4 dipilih. Misalkan X menandakan jumlah bola putih di dalam sampel. Bola tambahan sekarang dipilih dari enam bola selebihnya di dalam bejana. Misalkan Y sama dengan 1 jika bola berwarna putih dan 0 jika bola berwarna hitam. Dapatkan
(a) $E[Y | X = 2]$.
(b) $E[X | Y = 1]$.

(c) $\mathrm{Var}(Y \mid X = 0)$.

(e) $\mathrm{Var}(X \mid Y = 1)$.

35. Jika X dan Y adalah variabel acak eksponensial yang independen dan didistribusikan secara identik, perlihatkan bahwa distribusi bersyarat X, dengan $X + Y = t$, adalah distribusi seragam pada $(0, t)$.