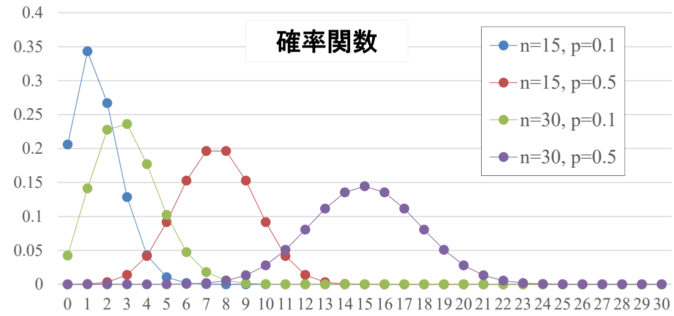
統計学入門 補助資料 ～代表的な確率分布～

2022年度1学期: 月曜2限

担当教員: 石垣 司

二項分布(離散分布)

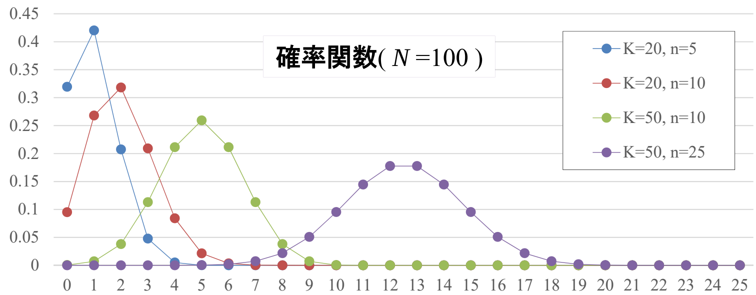
- n 回の独立なベルヌーイ試行の成功回数の分布
  - 例 10回のコイン投げで表が出る回数
  - 確率変数 $X = X_1 +, \cdots, +X_n$ （X_i は独立なベルヌーイ試行）
  - 確率関数 $\mathrm{Binomial}(X = x) = {}_nC_x p^x (1-p)^{n-x}$
  - パラメータ n, p



#メモ グラフの見やすさのために点と点の間に補間直線を引いているが、離散確率分布のため点（整数）以外の実数部分の値は意味を持たないことに注意

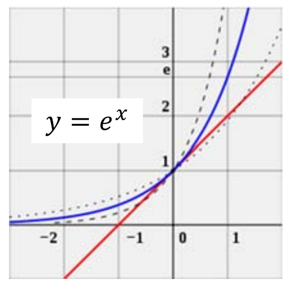
超幾何分布(離散分布)

- 有限母集団から非復元抽出の成功回数の分布
  - 例 N 個の製品の中に K 個の不良品が混入した。その中から n 個を取り出して梱包するとき，不良品の数が x 個である確率は？
  - 確率関数 $P(X = x) = \dfrac{{}_KC_x \times {}_{N-K}C_{n-x}}{{}_NC_n} = \dfrac{{}_nC_x \times {}_{N-n}C_{K-x}}{{}_NC_K}$
  - 平均と分散 $E[X] = \dfrac{nK}{N},\ V[X] = \dfrac{(N-n)n(N-K)K}{(N-1)N^2}$



補足：自然対数の底（ネイピア数）e

- 解析学(微分積分)で最も重要な数の一つ
  - $\dfrac{d}{dx} e^x = e^x$ となる定数 e をネイピア数とよぶ
  - $\int e^x \, dx = e^x + C$
  - $e \equiv \lim_{n \to \infty} \left(1 + \dfrac{1}{n}\right)^n \approx 2.71828 \cdots$（$e$ は無理数）
  - $\int \dfrac{1}{x} dx = \log_e x + C$ ：自然対数の底
  - 表記法：$e^x = \exp(x)$
  - 補足：オイラーの公式 $e^{i\pi} + 1 = 0$
    - 最も美しいと言われる数式の一つ



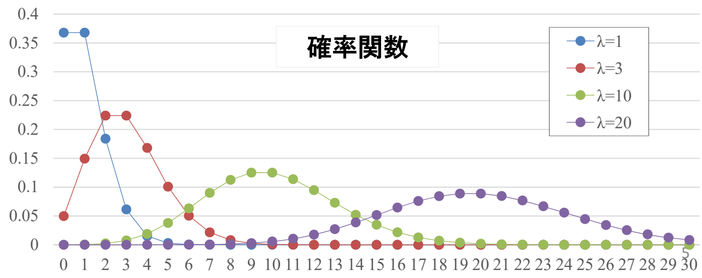
Wikipedia より (public domain)

ポアソン分布(離散分布)

Siméon Denis Poisson
(1781-1840)

- **稀にしか起きない事象の発生回数の分布**
  - **例　1日に起こる交通事故や倒産の回数**
  - **例　1時間以内の電話や来店客の回数**
  - **確率関数**　$\mathrm{Poisson}(X = x) = \frac{\lambda^x}{x!}\exp(-\lambda)$
  - **パラメータ**　λ（**単位時間当たりの平均発生回数**）
  - **平均と分散**　$E[X] = V[X] = \lambda$

※ポアソン分布の意義と使い方は単位時間(1秒，1時間，1日など)に依存する。単位時間の扱いに注意



演習問題

1. **二項分布の平均と分散を求めよ**
   ※**ベルヌーイ分布の平均と分散**（$E[X_i] = p, V[X_i] = p(1-p)$）
2. **20個(N)の製品の中に3個(K)の不良品が混入してしまった。その中から5個(n)を取り出して梱包するとき，不良品が1個以上混入している確率を求めよ**
3. **ある博物館には1時間に平均3人の来館者が来る。その傾向はポアソン分布に従うとする。このとき1時間に4人以上来館する確率を求めよ**
   (ここでは$\exp(-3) \cong 0.05$**として計算する)**

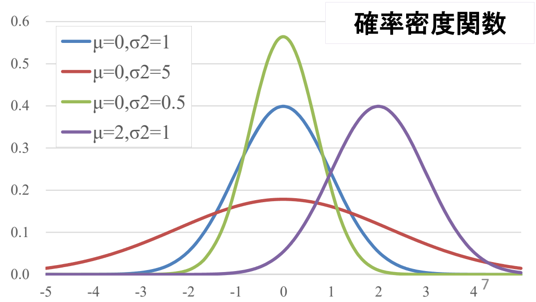
正規分布(連続分布)

Carolus Fridericus Gauss
(1777-1855)

- **統計学で最も重要な連続分布**
  - **実現象や観測誤差よく表現**
  - **確率密度関数**

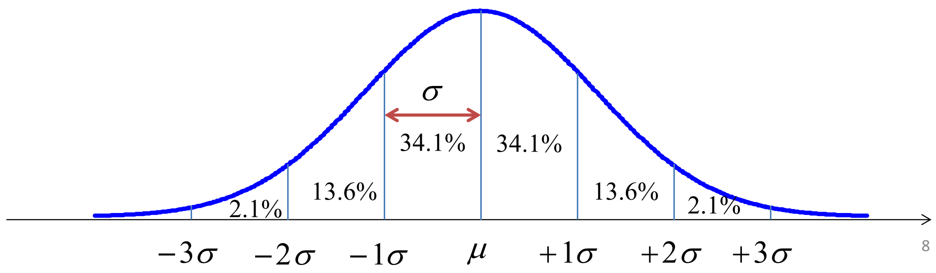
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

  - **パラメータ**　μ, σ^2
  - $E[X] = \mu, V[X] = \sigma^2$
  - $\mu = 0, \sigma^2 = 1$ **の正規分布を標準正規分布とよぶ**



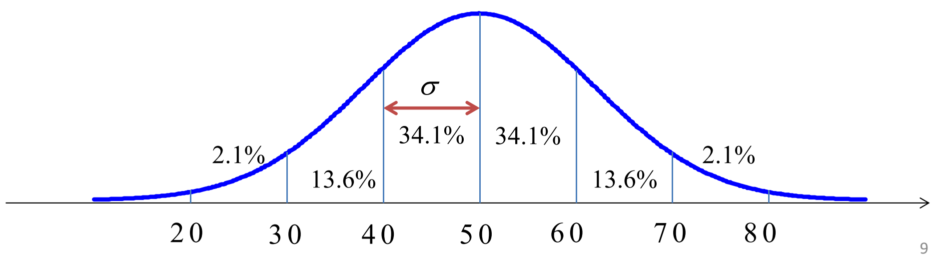
正規分布の確率密度関数の性質

- **左右対称の分布形**
  - **平均，中間値，最頻値は同じ**
  - **平均** μ **と分散** σ^2 **のみで分布が決まる(表記法：**$N(\mu, \sigma^2)$**)**
- **標準偏差** σ
  - $\pm\sigma$ **以内に事象が生じる確率 約**68.2%
    ($\pm 2\sigma$**は約**95.5%, $\pm 3\sigma$ **は約**99.7%)



正規分布を仮定する例

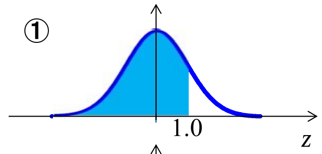
- **学力の偏差値**
  - **定義：平均**50, **標準偏差**10**の正規分布** $N(50,10^2)$
    - **偏差値**50 $\Rightarrow$ **成績上位**50%**くらいの人**
    - **偏差値**60 $\Rightarrow$ **成績上位**16%**くらいの人**
    - **偏差値**70 $\Rightarrow$ **成績上位**2%**くらいの人**
      ※絶対評価ではない。母集団の中での位置



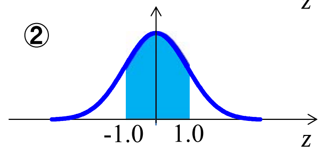
標準正規分布表

- **標準正規分布の各点の左側面積(下側確率)**
  - **縦軸：小数点第**1**位まで**
  - **横軸：小数点第**2**位**
- **問題**
  - **次の標準正規分布の青色部分の面積を求めよ**

①



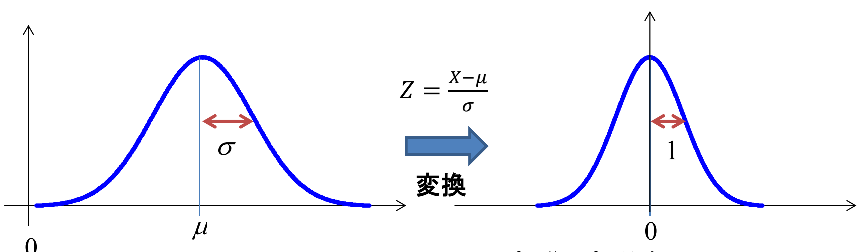
②



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

正規分布の標準化

- **確率変数** X **を** $X\sim N(\mu,\sigma^2)$ **とすると,**
  確率変数 $Z=\frac{X-\mu}{\sigma}$ **は,** $Z\sim N(0,1)$ **となる**
  - **正規分布は標準正規分布** $N(0,1)$ **に変換可能**
  - "~"**は, 確率変数** X **は確率分布** $N(\mu,\sigma^2)$ **から発生するという意味の記号**



正規分布 $X\sim N(\mu,\sigma^2)$　　標準正規分布　$Z\sim N(0,1)$

正規分布 $N(\mu,\sigma^2)$ の確率

- **問題**
  - **平均**2, **分散**9**の正規分布を考える。**

1. **その正規分布で** $P(X\le 3)$ **の値を求めよ**
2. **その正規分布で** $P(X\le x)=0.9$ **となる** x **の値を求めよ**

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

二項分布の正規近似

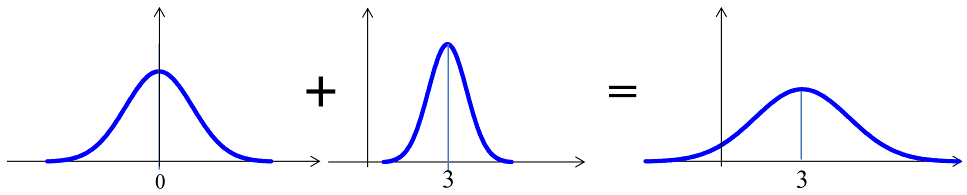
- **定理　二項分布に従う確率変数** $X \sim \mathrm{Binomial}(n,p)$ **は**, n **が十分に大きい時, 近似的に正規分布** $N(np, np(1-p))$ **に従う**
  - n **が大きい時, 2項分布の確率関数の計算は困難なため, 正規分布で近似した値が利用される**
  - **例: さいころを**700**回を投げて,** 1**の目の出現回数が**100**回以下となる確率を求めよ**
    - **二項分布を利用する場合,** $x = 0,1,\cdots,100$ **までの確率関数を計算**
    - **例:** $\mathrm{Binomial}(X=100) = {}_{700}C_{100}\left(\frac{1}{6}\right)^{100}\left\{1-\left(\frac{5}{6}\right)\right\}^{600}$
  - **正規近似の利用:** $X \sim N(np, np(1-p)) = N\left(\frac{700}{6}, \frac{700}{6}\frac{5}{6}\right)$ **の下で,** $P(X \le 100)$ **となる確率で近似可能**

正規分布の和の再生性

- **定理　互いに独立な確率変数** X_1 **と** X_2 **がそれぞれ** $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$**に従う。このとき、**$aX_1 + bX_2$ **の分布は** $N(a\mu_1 + b\mu_1, a^2\sigma_1^2 + b^2\sigma_2^2)$ **に従う**

  ※ **証明のためには積率母関数の知識が必要**

- **例:**



$X_1 \sim N(0,2)$　$X_2 \sim N(3,1)$　$X_1 + X_2 \sim N(3,3)$

補足:その他の分布の再生性

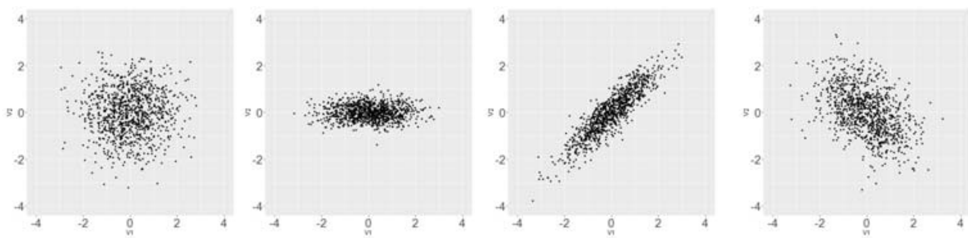
- **再生性をもつ代表的な分布**
  - **正規分布**
    $X_i \sim N(\mu_i, \sigma_i^2) \Rightarrow X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$
  - **ガンマ分布**
    $X_i \sim \mathrm{Gamma}(k_i, \theta) \Rightarrow X_1 + X_2 \sim \mathrm{Gamma}(k_1 + k_2, \theta)$
  - **二項分布**
    $X_i \sim \mathrm{Binomial}(n_i, p) \Rightarrow X_1 + X_2 \sim \mathrm{Binomial}(n_1 + n_2, p)$
  - **ポアソン分布**
    $X_i \sim \mathrm{Poisson}(\lambda_i) \Rightarrow X_1 + X_2 \sim \mathrm{Poisson}(\lambda_1 + \lambda_2)$

補足:2変量正規分布

- **確率変数** X **と** Y **に関する正規分布**
  - $E[X], E[Y], V[X], V[Y], Cov[X,Y]$ **で分布が決まる**

    ※**学部上級・大学院以上の計量経済学では多変量扱いが必須**

$V[X]=1$	$V[X]=1$	$V[X]=1$	$V[X]=1$
$V[Y]=1$	$V[Y]=0.1$	$V[Y]=1$	$V[Y]=0.5$
$Cov[X,Y]=0$	$Cov[X,Y]=0$	$Cov[X,Y]=0.9$	$Cov[X,Y]=-0.5$



各代表値を持つ2変量正規分布から1000**個の乱数を発生**

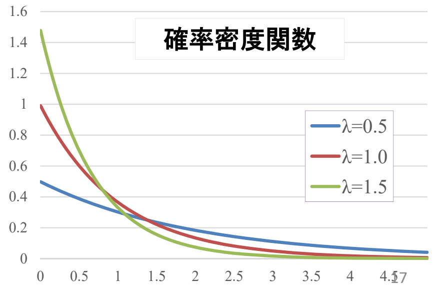
補足:指数分布(連続分布)

- ポアソン分布(パラメータ λ)に従う事象が1回生じるまでの期間の分布
    - ポアソン分布:単位時間内での事象の発生回数の分布
  - 確率密度関数

$$f(x) = \lambda \exp(-\lambda x), (x \geq 0)$$

  - パラメータ　$\lambda > 0$
  - 平均と分散

$E[X] = \lambda^{-1}, V[X] = \lambda^{-2}$



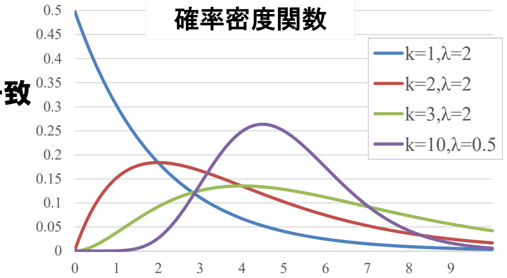
補足:ガンマ分布(連続分布)

- パラメータ λ の指数分布に従う事象が k 回起こるまでの確率分布
  - 確率密度関数　$f(x) = \frac{\lambda^k}{\Gamma(k)} x^{k-1} \exp(-\lambda x), (x \geq 0)$
    - $\Gamma(k)$はガンマ関数
  - パラメータ

$k > 0, \lambda^{-1} > 0$

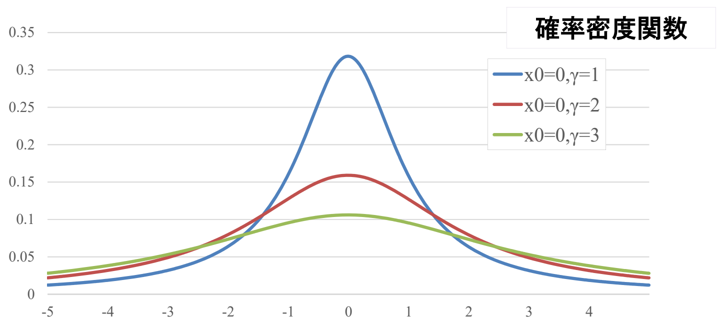
    - $k = 0$ のとき指数分布に一致
  - 平均と分散

$E[X] = k\lambda^{-1}, V[X] = k\lambda^{-2}$



補足:コーシー分布(連続分布)

- 平均も分散も存在しない分布
  - 確率密度関数　$f(x) = \frac{\gamma}{\pi\{\gamma^2 + (x - x_0)^2\}}$
    - 期待値の計算が不定・発散(平均は定まらない。分散は∞)
    - 正規分布に似ているが両裾が厚い



演習問題

- 米国女性の妊娠から出産までの期間は平均266日
- 妊娠期間は標準偏差16日の正規分布にきれいに従う
- Aさんは米国人で, 夫は海軍の船乗りである
- 「私は正確に310日間妊娠し, 子供を出産しました。なぜなら, 妊娠した日から子供が生まれるまで夫には1回も会わなかったからです」

- 問題
  1. 標準正規分布表から米国女性が310日間以上妊娠する確率を求めよ
  2. その確率を聞いて, Aさんは浮気を疑われることを恐れた。統計的な観点からAさんを安心させよ

（「T.H.ダベンポート、キム・ジノ、真実を見抜く分析力、日経BP、2014」より変更し作問）