確率分布 2

連続型の確率分布の導出を行います。導出と平均を求めているだけです。
指数分布、ガンマ分布、χ^2 分布、F-分布、t-分布を扱っています。
基本的に確率変数は大文字、その値はそれの小文字で書きます。

連続的な確率変数の場合での定義を与えておきます。
先に離散的なときも見ておきます。離散的なときの確率変数 X は離散的な値 $x_1, x_2, \ldots, x_n$ を持ち、それぞれの確率は $P(X = x_i)$ のように書かれます。離散的な場合は、確率はそのまま確率分布 f になっていて

$$P(X = x_i) = f(x_i)\ ,\ \sum_{i=1}^{n} f(x_i) = 1\ ,\ f(x_i) \geq 0$$

累積分布関数は

$$P(X \leq x_m) = \sum_{i=1}^{m} f(x_i)$$

と定義されます。
2 つの離散的な確率変数 X, Y があったとき、$X = x, Y = y$ となる確率は $P(x, y)$ と書かれ、同時確率質量関数 (joint probability mass function) と呼ばれます。この X, Y の組み合わせが A となる場合

$$P(A) = \sum_{x,y \in A} P(x, y)$$

のようになります。例えば、$x + y = z$ なら、$x + y$ が z になるように和を取ります。
X, Y が独立であるなら

$$P(x, y) = P_x(x) P_y(y)$$

のようにそれぞれの確率の積で書くことが出来ます。さらに複数の確率変数があっても同様です。例えば、1 から 10 までの数字があり、それぞれが出力される確率は $P_i\ (i = 1, 2, \ldots, 10)$ とします。2 回出力させて、m, n が出てくる確率は

$$P = P_m P_n$$

と書けるということです。また、これに例えば数字の和が 10 となるという制限があるなら、10 となる組み合わせの確率を足し合わせるので

$$P(10) = \sum_{i,j} P_i P_j \quad (i + j = 10)$$

となります。
確率変数 X の値が連続的であるなら、確率変数は連続的と言われます (より細かい言い回しもありますが、省きます)。連続的なので添え字を省いて x とします。このとき確率分布 (確率密度関数)$f(x)$ は、a から b までを合わせた確率 $P(a \leq X \leq b)$ を

$$P(a \leq X \leq b) = \int_a^b dx\ f(x) \quad (a < b\ ,\ f(x) \geq 0)$$

と与えるものとして定義され、$f(x)dx$ は x から $x+dx$ の間の確率を与えます。連続的な場合では、確率分布は確率そのものになっていません。規格化は

$$\int dx\ f(x) = 1$$

となります。積分範囲は x の可能な範囲です。もしくは $-\infty \sim +\infty$ として、$f(x)$ がどこかの領域から 0 になるとしてもいいです。

累積分布関数 $P(X \leq x)$ は確率分布 $f(x)$ から

$$P(X \leq x) = \int_{-\infty}^{x} dx'\ f(x')$$

と定義されます。これは X が $X \leq x$ となる確率です。これから確率分布は

$$\frac{d}{dx}P(X \leq x) = \frac{d}{dx}\int_{-\infty}^{x} dx'\ f(x') = f(x)$$

と求まります。

確率変数 X, Y があるとき、その確率は

$$P(A) = \int_A dxdy\ f(x,y) \quad (\int_{-\infty}^{\infty} dxdy\ f(x,y) = 1)$$

として与えられます。添え字の A は x, y の取る範囲に対してで、$f(x,y)dxdy$ は、x が x から $x+dx$、y が y から $y+dx$ となる確率を与えます。$f(x,y)$ を同時確率密度関数 (joint probability density function) と呼びますが、ここでは区別せずに確率分布と言ってしまいます。

X, Y が独立なら $f(x,y) = f_1(x)f_2(y)$ となり

$$P(A) = \int_A dxdy\ f_1(x)f_2(y)$$

と書けます。f_1, f_2 は確率変数 X, Y が従う確率分布です。

確率変数 X_1, X_2 があるとし、値は $-\infty$ から $+\infty$ まで取れるとします。確率は

$$P(A) = \int_A dx_1dx_2\ g(x_1, x_2)$$

で与えます。このとき、$Y = X_1 + X_2$ として、$Y \leq y$ となる累積分布関数は

$$P(Y \leq y) = P(X_1 + X_2 \leq y) = \int_{x_1+x_2 \leq y} dx_1dx_2\ g(x_1, x_2)$$

となります。そうすると、$X_2 \leq y - X_1$ なので

$$P(Y \leq y) = \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{y-x_1} dx_2\ g(x_1, x_2)$$

と書けて、y 微分は

$$f(y) = \frac{d}{dy}P(Y \leq y) = \int_{-\infty}^{\infty} dx_1\ g(x_1, y - x_1) \tag{1}$$

これは Y が従う確率分布となります。そして、この形はデルタ関数を使って

$$f(y) = \int_{-\infty}^{\infty} dx_1 dx_2 \; g(x_1, x_2)\delta(y - (x_1 + x_2)) \tag{2}$$

と書くことが出来ます。一般化して、$Y = h(X_1, X_2)$ とすれば

$$f(y) = \int_{-\infty}^{\infty} dx_1 dx_2 \; g(x_1, x_2)\delta(y - h(x_1, x_2))$$

となります。さらに、$Y_1 = u(X_1, X_2), Y_2 = v(X_1, X_2)$ とすれば、Y_1, Y_2 が従う確率分布 $f(y_1, y_2)$ は

$$f(y_1, y_2) = \int_{-\infty}^{\infty} dx_1 dx_2 \; g(x_1, x_2)\delta(y_1 - u(x_1, x_2))\delta(y_2 - v(x_1, x_2))$$

と書けます。このように確率変数の変換はデルタ関数によって行えます。
　ついでに、ガンマ関数とベータ関数が出てくるので、定義を与えておくと、ガンマ関数は

$$\Gamma(\alpha) = \int_0^{\infty} dx \; x^{\alpha-1} e^{-x} \quad (\alpha > 0)$$

ベータ関数は

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} \; , \; B(\alpha, \beta) = \int_0^1 dx \; x^{\alpha-1}(1 - x)^{\beta-1}$$

- 正規分布

  ガウス分布のことですが、ここでは確率や統計よりの話なので、正規分布 (normal distribution) と呼んでいきます。導出は「ガウス分布」で行っているので飛ばします (ここでは記号の定義を変えています)。正規分布は

  $$N(x; \mu, \sigma) = C \exp[-\frac{1}{2\sigma}(x - \mu)^2] \quad (-\infty < x < +\infty)$$

  と定義されています。C は規格化定数で

  $$C = \sqrt{\frac{1}{2\pi\sigma}}$$

  となっています。確率変数 X (x は X の値) が正規分布に従っています。μ は x の平均、σ は x の分散です。$\mu = 0$ と $\sigma = 1$ にした

  $$N(x) = \frac{1}{\sqrt{2\pi}} \exp[-\frac{1}{2}x^2]$$

  を標準正規分布 (standard normal distribution) と言います。

- 指数分布

　ポアソン分布から指数分布を求めます。ポアソン分布は時間間隔 t で適当な事象が k 回発生する確率を与え、

$$P(k,\alpha) = \frac{(\alpha t)^k}{k!}e^{-\alpha t} \quad (\alpha > 0)$$

となっています。

　ポアソン分布は変数として k を使っていますが、もう一つ外部から与えられているものとして時間間隔があります。このことから、ある時間間隔 t で事象が起こる確率分布に変更できると考えられます。つまり、事象と事象の時間間隔を確率変数 T とします。

　まず、時間間隔 t で 1 回も発生しないときの確率は

$$P(0,\alpha) = e^{-\alpha t} \quad (k = 0)$$

と与えられます。

　ここで、累積分布関数を持ち出します。累積分布関数 $P(T \le t)$ は確率分布 $f(t)$ から

$$P(T \le t) = \int_{-\infty}^{t} dt'\ f(t')$$

とします (T が t 以下のときでの確率)。今の場合、T は事象と事象の間の時間間隔なので、$T \le t$ では事象が起きることから、$P(T \le t)$ は時間間隔 t で事象が起きる確率となります (実際には時間間隔は負の値を持たないので下限は 0)。ここで

$$1 = \int_{-\infty}^{\infty} dt'\ P(t') = \int_{-\infty}^{t} dt'\ P(t') + \int_{t}^{\infty} dt'\ P(t')$$

と分解してみます。第一項は累積分布 $P(T \le t)$ です。変形すれば

$$\int_{t}^{\infty} dt'\ P(t') = 1 - \int_{-\infty}^{t} dt'\ P(t') = 1 - P(T \le t)$$

なので、左辺は $P(T \le t)$ でない確率を表すことになります。そうすると、$P(T \le t)$ は時間間隔 t で事象が起きる確率なので、左辺は時間間隔 t で事象が起きない確率となります。これを $P(T > t)$ と書くことにします。$P(T > t)$ は tail distribution や complementary cumulative distribution と呼ばれます。

　時間間隔 t で事象が起きないことは、ポアソン分布での $k = 0$ のことなので、

$$P(T > t) = e^{-\alpha t}$$

となります。よって

$$P(T \le t) = 1 - P(T > t) = 1 - e^{-\alpha t}$$

確率分布 $f(t)$ はこれを t で微分すればいいので

$$f(t) = \frac{d}{dt}(1 - e^{-\alpha t}) = \alpha e^{-\alpha t} \qquad (t > 0)$$

となります。これを指数分布 (exponential distribution) と呼びます。導出から分かるように、ポアソン分布を満たす事象が起きる時間間隔 t に対する確率分布を与えています。言い換えれば、事象が起きるまでの待機時間 t が従う確率分布です。

平均 $< t >$ は部分積分から

$$< t >= \alpha \int_0^\infty dt\ te^{-\alpha t} = \int_0^\infty dt\ e^{-\alpha t} = \frac{1}{\alpha}$$

と分かります。t^2 でもすぐに

$$< t^2 >= \alpha \int_0^\infty dt\ t^2 e^{-\alpha t} = \alpha[-\frac{1}{\alpha} t^2 e^{-\alpha t}] + 2\int_0^\infty dt\ te^{-\alpha t} = \frac{2}{\alpha}$$

よって、分散は

$$< t^2 > - < t >^2 = \frac{1}{\alpha^2}$$

となります。

- ガンマ分布

　指数分布から少し状況を変えます。指数分布では事象が起きるかどうかでしたが、今度は r 番目の事象が起きるまでの時間間隔に対する確率を求めます。導出は指数関数と同じで、まず、累積分布関数 $P(T \leq t)$ を用意して

$$P(T \leq t) = 1 - P(T > t)$$

とします。$P(T \leq t)$ は時間間隔 t で事象が r 回起きる確率 (r 番目の事象が起きる確率)、$P(T > t)$ は t で r 回起きない確率です。

　ポアソン分布は時間間隔 t で k 回起きる確率なので、$k = 0$ から $k = r - 1$ までの和は 0 回から $r - 1$ 回起きたときのそれぞれの確率の和です。つまり、r 回起きない確率になります。そうすると、$P(T > t)$ はポアソン分布から

$$P(T > t) = \sum_{k=0}^{r-1} \frac{(\alpha t)^k}{k!} e^{-\alpha t}$$

なので

$$P(T \leq t) = 1 - \sum_{k=0}^{r-1} \frac{(\alpha t)^k}{k!} e^{-\alpha t}$$

後は t で微分して

$$
\begin{aligned}
f(t,r) &= \frac{d}{dt}(1-\sum_{k=0}^{r-1}\frac{(\alpha t)^k}{k!}e^{-\alpha t}) \\
&= -\sum_{k=0}^{r-1}\frac{k\alpha^k t^{k-1}}{k!}e^{-\alpha t}+\alpha\sum_{k=0}^{r-1}\frac{(\alpha t)^k}{k!}e^{-\alpha t} \\
&= -\sum_{k=1}^{r-1}\frac{\alpha^k t^{k-1}}{(k-1)!}e^{-\alpha t}+\alpha\sum_{k=0}^{r-1}\frac{(\alpha t)^k}{k!}e^{-\alpha t} \\
&= -\alpha\sum_{k=1}^{r-1}\frac{(\alpha t)^{k-1}}{(k-1)!}e^{-\alpha t}+\alpha\sum_{k=0}^{r-1}\frac{(\alpha t)^k}{k!}e^{-\alpha t} \\
&= -\alpha\sum_{n=0}^{r-2}\frac{(\alpha t)^{n}}{n!}e^{-\alpha t}+\alpha\sum_{k=0}^{r-1}\frac{(\alpha t)^k}{k!}e^{-\alpha t}\qquad (n=k-1) \\
&= -\alpha\sum_{n=0}^{r-2}\frac{(\alpha t)^{n}}{n!}e^{-\alpha t}+\alpha\sum_{k=0}^{r-2}\frac{(\alpha t)^k}{k!}e^{-\alpha t}+\alpha\frac{(\alpha t)^{(r-1)}}{(r-1)!}e^{-\alpha t} \\
&= \alpha\frac{(\alpha t)^{(r-1)}}{(r-1)!}e^{-\alpha t} \\
&= \frac{\alpha^r}{(r-1)!}t^{r-1}e^{-\alpha t}
\end{aligned}
$$

よって、時間間隔 t で r 回の事象が起きる確率を与える (r 回起きるまでの待機時間が従う) 確率分布は

$$
f(t,r)=\frac{\alpha^r}{(r-1)!}t^{r-1}e^{-\alpha t}=\frac{\alpha^r}{\Gamma(r)}t^{r-1}e^{-\alpha t}\qquad (t>0\ ,\ \alpha>0)
$$

となり、これをガンマ分布 (gamma distribution) と呼びます。r は正の実数、$\Gamma(r)$ はガンマ関数です。
　ちなみに、これの t 積分は

$$
\begin{aligned}
\int_0^\infty dt\ t^{r-1}e^{-\alpha t} &= \frac{1}{\alpha}\int_0^\infty dt'\ (\frac{t'}{\alpha})^{r-1}e^{-t'}\qquad (t'=\alpha t\ ,\ dt'=\alpha dt) \\
&= \frac{1}{\alpha^r}\int_0^\infty dt\ t^{r-1}e^{-t} \\
&= \frac{1}{\alpha^r}\Gamma(r)
\end{aligned}
$$

となります。これからガンマ分布 $f(t,r)$ は 1 に規格化されているのも確認できます。
　$r=1$ では時間間隔 t で 1 回の事象が起きる確率になるので、

$$
P(t,r=1)=\alpha e^{-\alpha t}
$$

となって、指数分布になります。
　平均 $<t>$ は

$$
\begin{aligned}
< t >&= \frac{\alpha^r}{\Gamma(r)} \int_0^\infty dt\ t t^{r-1} e^{-\alpha t} \\
&= \frac{\alpha^r}{\Gamma(r)} \frac{1}{\alpha} \int_0^\infty dx\ \frac{1}{\alpha} (\frac{1}{\alpha})^{r-1} x^{r+1-1} e^{-x} \qquad (x = \alpha t) \\
&= \frac{1}{\alpha} \frac{1}{\Gamma(r)} \int_0^\infty dx\ x^{r+1-1} e^{-x} \\
&= \frac{1}{\alpha} \frac{1}{\Gamma(r)} \Gamma(r+1)
\end{aligned}
$$

これはそのまま一般化できて

$$
\begin{aligned}
< t^k >&= \frac{\alpha^r}{\Gamma(r)} \int dt\ t^{k+r-1} e^{-\alpha t} \\
&= \frac{\alpha^r}{\Gamma(r)} \frac{1}{\alpha} \int dx\ (\frac{1}{\alpha})^{k+r-1} x^{r+1-1} e^{-x} \\
&= (\frac{1}{\alpha})^k \frac{1}{\Gamma(r)} \int dx\ x^{k+r-1} e^{-x} \\
&= (\frac{1}{\alpha})^k \frac{1}{\Gamma(r)} \Gamma(k+r)
\end{aligned}
$$

となります。

- χ^2 分布 (カイ二乗分布)
  n 個の独立な確率変数 X_i $(i = 1, 2, \ldots, X_n)$ が標準正規分布

$$
N(x_i) = \frac{1}{\sqrt{2\pi}} \exp[-\frac{1}{2} x_i^2] \qquad (-\infty < x_i < \infty)
$$

に従っているとします。そして、これらによって作られる確率変数 Z として

$$
Z = \sum_{i=1}^n X_i^2 \qquad (z = \sum_{i=1}^n x_i^2)
$$

というのを考え、Z が従う確率分布を求めます。

まず、$n = 1$ とし、$Z = X_1^2 = Y_1$ とします。X_1 が従うのは標準正規分布です。これから、x_1 を $z = x_1^2$ に書き換えます。つまり、変数変換を行います。

積分形にしたほうが分かりやすいので、累積分布関数で行います。$X_1^2 \leq u$ での累積分布関数は X_1 が標準正規分布に従っていることから

$$
\begin{aligned}
P_1(X_1^2 \le u) &= P(-\sqrt{u} \le X_1 \le \sqrt{u}) \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{u}}^{\sqrt{u}} dx_1 \ \exp[-\frac{1}{2}x_1^2] \\
&= \frac{2}{\sqrt{2\pi}} \int_{0}^{\sqrt{u}} dx_1 \ \exp[-\frac{1}{2}x_1^2] \\
&= \frac{1}{\sqrt{2\pi}} \int_{0}^{u} dy_1 \ \frac{1}{\sqrt{y_1}} \exp[-\frac{1}{2}y_1] \quad (y_1 = x_1^2 \ dy_1 = 2x_1 dx_1)
\end{aligned}
$$

$y_1 = x_1^2$ なので y_1 の下限は 0 です。これを u で微分すれば確率分布になるので

$$
\begin{aligned}
\frac{d}{du}P_1(X_1^2 \le u) &= \frac{1}{\sqrt{2\pi}} \frac{d}{du} \int_{0}^{u} dy_1 \ \frac{1}{\sqrt{y_1}} \exp[-\frac{1}{2}y_1] \\
&= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{u}} \exp[-\frac{1}{2}u]
\end{aligned}
$$

よって、$Z = X_1^2$ が従う確率分布は

$$
f_1(z) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{z}} \exp[-\frac{1}{2}z]
$$

となります。

今度は $Z = X_1^2 + X_2^2$ とします。このときは 2 つの確率変数 X_1, X_2 があり、それぞれが独立に標準正規分布に従っています。なので、(1) を使います。

X_1, X_2 での確率分布 $f(x_1, x_2)$ は 2 つの標準正規分布の積 $N(x_1)N(x_2)$ です。このときの累積分布関数は

$$
P_2(Z \le u) = P_2(X_1^2 + X_2^2 \le u) = \int_{x_1^2 + x_2^2 \le u} dx_1 dx_2 \ N(x_1)N(x_2)
$$

と書けます。これの変数を $y_1 = x_1^2, y_2 = x_2^2$ に変えます。$X_1^2 + X_2^2 \le u$ は $X_2^2 \le u - X_1^2$ なので

$$
\begin{aligned}
P_2(X_1^2+X_2^2 \le u) &= P_2(X_2^2 \le u - X_1^2) \\
&= P_2(-\sqrt{u-X_1^2} \le X_2 \le \sqrt{u-X_1^2}) \\
&= \int_{-\infty}^{\infty} dx_1 \int_{-v}^{+v} dx_2 \ N(x_1)N(x_2) \quad (v=\sqrt{u-x_1^2}) \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} dx_1 \int_{-v}^{+v} dx_2 \ \exp[-\frac{1}{2}x_1^2]\exp[-\frac{1}{2}x_2^2] \\
&= \frac{2}{2\pi}\int_{-\infty}^{\infty} dx_1 \exp[-\frac{1}{2}x_1^2]\int_0^{v} dx_2 \exp[-\frac{1}{2}x_2^2] \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} dx_1 \exp[-\frac{1}{2}x_1^2]\int_0^{u-x_1^2} dy_2 \frac{1}{\sqrt{y_2}}\exp[-\frac{1}{2}y_2]
\end{aligned}
$$

u で微分して、$Z = X_1^2 + X_2^2$ が従う確率分布 $f_2(u)$ にすると

$$
\begin{aligned}
f_2(u) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} dx_1 \frac{1}{\sqrt{u-x_1^2}}\exp[-\frac{1}{2}x_1^2]\exp[-\frac{1}{2}(u-x_1^2)] \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} dx_1 \frac{1}{\sqrt{u-x_1^2}}\exp[-\frac{1}{2}u)] \\
&= \frac{2}{2\pi}\int_{0}^{\infty} dx_1 \frac{1}{\sqrt{u-x_1^2}}\exp[-\frac{1}{2}u)] \\
&= \frac{1}{2\pi}\int_{0}^{\infty} dy_1 \frac{1}{\sqrt{y_1}}\frac{1}{\sqrt{u-y_1}}\exp[-\frac{1}{2}u)]
\end{aligned}
$$

y_1 が u を超えると虚数が出てくるので上限を u にして (元々の制限 $X_1^2 + X_2^2 \le u$ からも)

$$
f_2(u) = \frac{1}{2\pi}\int_0^{u} dy_1 \frac{1}{\sqrt{y_1}}\frac{1}{\sqrt{u-y_1}}\exp[-\frac{1}{2}u]
$$

これの積分は

$$
\begin{aligned}
f_2(z) &= \frac{1}{2\pi}\exp[-\frac{1}{2}z]\int_0^{z} dy \ \frac{1}{\sqrt{y(z-y)}} \\
&= \frac{1}{2\pi}\exp[-\frac{1}{2}z]\int_0^{z} dy \ \frac{1}{\sqrt{yz(1-y/z)}} \\
&= \frac{1}{2\pi}\exp[-\frac{1}{2}z]\int_0^{1} d\alpha \ \frac{z}{\sqrt{yz}\sqrt{(1-\alpha)}} \quad (\alpha = \frac{y}{z} \ , \ d\alpha = \frac{1}{z}dy) \\
&= \frac{1}{2\pi}\exp[-\frac{1}{2}z]\int_0^{1} d\alpha \ \sqrt{\frac{z}{y}}\frac{1}{\sqrt{(1-\alpha)}} \\
&= \frac{1}{2\pi}\exp[-\frac{1}{2}z]\int_0^{1} d\alpha \ \frac{1}{\sqrt{\alpha(1-\alpha)}}
\end{aligned}
$$

面倒なので積分の結果として

$$\int d\alpha\ \frac{1}{\sqrt{-\alpha(\alpha-1)}} = \arcsin(2\alpha-1) \qquad (0<\alpha<1)$$

を使えば

$$\begin{aligned}
f_2(z) &= \frac{1}{2\pi}\exp[-\frac{1}{2}z]\int_0^1 d\alpha\ \frac{1}{\sqrt{\alpha(1-\alpha)}} \\
&= \frac{1}{2\pi}\exp[-\frac{1}{2}z](\arcsin(1)-\arcsin(-1)) \\
&= \frac{1}{\pi}\exp[-\frac{1}{2}z]\arcsin(1) \qquad (\arcsin(\theta) = -\arcsin(-\theta))
\end{aligned}$$

逆三角関数の定義から

$$\arcsin(1) = \frac{\pi}{2}\ \Leftrightarrow\ \sin\frac{\pi}{2} = 1$$

なので

$$P_2(z) = \frac{1}{2}\exp[-\frac{1}{2}z]$$

となります。

同じ結果を一般化しやすい形から求められます。(2) から、デルタ関数を使えば $f_2(z)$ は

$$f_2(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dx_1dx_2\ N(x_1)N(x_2)\delta(z-(x_1^2+x_2^2))$$

と書けます。dx_1dx_2 を極座標にして

$$rdrd\theta = rdrdS_2$$

とすれば

$$\begin{aligned}
f_2(z) &= \frac{1}{2\pi}\int_0^{\infty} dr\ r\int dS_2\ \exp[-\frac{1}{2}r^2]\delta(z-r^2) \\
&= \frac{1}{2\pi}\int_0^{\infty} dr\ r\int_0^{2\pi} d\theta\ \exp[-\frac{1}{2}r^2]\delta(z-r^2) \\
&= \frac{1}{2\pi}\int_0^{\infty} dr\ 2\pi r\exp[-\frac{1}{2}r^2]\delta(z-r^2) \\
&= \int_0^{\infty} d\rho\ \frac{1}{2}\exp[-\frac{1}{2}\rho]\delta(z-\rho) \qquad (\rho = r^2\ ,\ d\rho = 2rdr) \\
&= \frac{1}{2}\exp[-\frac{1}{2}z]
\end{aligned}$$

となって、同じ結果になります。S_2 は半径 1 での表面積 (円周) です。

そうすると、確率変数を n 個に一般化したとき

$$f_n(z) = \frac{1}{(2\pi)^{n/2}} \int_{-\infty}^{\infty} dx_1 dx_2 \cdots dx_n \ \exp[-\frac{1}{2}(x_1^2 + x_2^2 + \cdots + x_n^2)]\delta(z - (x_1^2 + x_2^2 + \cdots + x_n^2))$$

同様に変形していけば

$$f_n(z) = \frac{1}{(2\pi)^{n/2}} \int_0^{\infty} dr \ r^{n-1} \int dS_n \ \exp[-\frac{1}{2}r^2]\delta(z - r^2) \quad (r^2 = x_1^2 + x_2^2 + \cdots + x_n^2)$$

n 次元の半径 1 の球の表面積は

$$S_n = \frac{n\pi^{n/2}}{\Gamma(n/2+1)}$$

となっています。$\Gamma(\alpha)$ はガンマ関数です。これからすぐに

$$\begin{aligned}
f_n(z) &= \frac{1}{(2\pi)^{n/2}} \frac{1}{2} z^{n/2-1} \exp[-\frac{1}{2}z] \int dS_n \\
&= \frac{1}{(2\pi)^{n/2}} \frac{1}{2} z^{n/2-1} e^{-z/2} \frac{n\pi^{n/2}}{\Gamma(n/2+1)} \\
&= \frac{n}{2^{n/2}} \frac{1}{2} \frac{z^{n/2-1} e^{-z/2}}{\Gamma(n/2+1)} \\
&= \frac{n}{2^{n/2}} \frac{1}{2} \frac{2}{n} \frac{z^{n/2-1} e^{-z/2}}{\Gamma(n/2)} \quad (\Gamma(\alpha+1) = \alpha\Gamma(\alpha)) \\
&= \frac{1}{2^{n/2}} \frac{z^{n/2-1} e^{-z/2}}{\Gamma(n/2)} \quad (z \geq 0)
\end{aligned}$$

となります。この、標準正規分布に従う確率変数 X_i による確率変数 $Z = X_1^2 + \cdots + X_n^2$ が従う確率分布 $f_n(z)$ を、χ^2 分布 (カイ 2 乗分布、chi-squared distribution) と呼びます

z^k の平均 $< z^k >$ を求めます。$< z >$ は

$$\begin{aligned}
< z > &= \frac{1}{2^{n/2}\Gamma(n/2)} \int_0^{\infty} dz \ z^{n/2-1} e^{-z/2} \\
&= \frac{2}{2^{n/2}\Gamma(n/2)} \int_0^{\infty} dx \ 2^{n/2} x^{n/2-1+1} e^{-x} \quad (x = \frac{z}{2})
\end{aligned}$$

これはガンマ関数

$$\Gamma(\alpha) = \int_0^{\infty} dx \ x^{\alpha-1} e^{-x} \quad (\alpha > 0)$$

を使えば

$$\int_0^\infty dx\ x^{n/2-1+1}e^{-x} = \Gamma(\frac{n}{2}+1)$$

と書けます。よって

$$< z >= \frac{2}{\Gamma(n/2)}\Gamma(\frac{n}{2}+1)$$

これはすぐに z^k に一般化できることが分かって

$$< z^k >= \frac{2^k}{\Gamma(n/2)}\Gamma(\frac{n}{2}+k)$$

見て分かるように、実際の計算ではガンマ関数の値が必要になります。これを使えば分散 $< z^2 > - < z >^2$ も求まります。

- F-分布

  独立な確率変数 X, Y を用意し、それぞれが χ^2 分布に従っているとします。なので、それぞれの確率分布は

$$\chi(x,m) = \frac{1}{2^{m/2}}\frac{x^{m/2-1}e^{-x/2}}{\Gamma(m/2)}\ ,\ \chi(y,n) = \frac{1}{2^{n/2}}\frac{y^{n/2-1}e^{-y/2}}{\Gamma(m/2)}\quad (x, y \geq 0)$$

独立にしているので、合わせた確率は $\chi(x,m)\chi(y,n)$ を積分すればいいです。これに対して、X, Y の比率 X/Y の累積分布関数を考えます。$X/Y \leq t$ は $X \leq tY$ なので、累積分布関数は

$$P(\frac{X}{Y} \leq t) = P(X \leq tY) = \int_0^\infty dy \int_0^{ty} dx\ \chi(x,m)\chi(y,n)$$

これを t 微分して

$$\begin{aligned}
\frac{d}{dt}P(\frac{X}{Y} \leq t) &= \frac{d}{dt}\int_0^\infty dy \int_0^{ty} dx\ \chi(x,m)\chi(y,n) \\
&= \int_0^\infty dy\ y\frac{d}{dt'}f\int_0^{t'} dx\ \chi(x,m)\chi(y,n) \quad (t' = ty) \\
&= \int_0^\infty dy\ y\chi(ty,m)\chi(y,n) \\
&= \int_0^\infty dy\ y\frac{1}{2^{m/2}}\frac{(ty)^{m/2-1}e^{-ty/2}}{\Gamma(m/2)}\frac{1}{2^{n/2}}\frac{y^{n/2-1}e^{-y/2}}{\Gamma(n/2)} \\
&= (\frac{1}{2})^{(m+n)/2}\frac{1}{\Gamma(m/2)\Gamma(n/2)}t^{m/2-1}\int_0^\infty dy\ y^{(m+n)/2-1}e^{-(t+1)y/2}
\end{aligned}$$

積分はガンマ関数を利用すると実行できます。ガンマ関数は

$$
\begin{aligned}
\Gamma(n) &= \int_0^\infty dx\ x^{n-1}e^{-x} \\
&= \int_0^\infty dy\ \beta(\beta y)^{n-1}x^{n-1}e^{-\beta y} \qquad (x=\beta y) \\
&= \int_0^\infty dy\ \beta^n y^{n-1}e^{-\beta y}
\end{aligned}
$$

と書き換えられるので

$$
\begin{aligned}
\int_0^\infty dy\ y^{(m+n)/2-1}e^{-(t+1)y/2} &= \int_0^\infty dy\ y^{\alpha-1}e^{-\beta y} \qquad (\alpha=\frac{m+n}{2}\ ,\ \beta=\frac{t+1}{2}) \\
&= \frac{1}{\beta^\alpha}\int_0^\infty dy\ \beta^\alpha y^{\alpha-1}e^{-\beta y} \\
&= \frac{1}{\beta^\alpha}\Gamma(\alpha)
\end{aligned}
$$

これから

$$
\begin{aligned}
\frac{d}{dt}P(\frac{X}{Y}\leq t) &= (\frac{1}{2})^{(m+n)/2}\frac{1}{\Gamma(m/2)\Gamma(n/2)}t^{m/2-1}\frac{1}{\beta^\alpha}\Gamma(\alpha,\beta) \\
&= (\frac{1}{2})^{(m+n)/2}\frac{1}{\Gamma(m/2)\Gamma(n/2)}t^{m/2-1}(\frac{2}{t+1})^{(m+n)/2}\Gamma(\frac{m+n}{2}) \\
&= \frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)}t^{m/2-1}(t+1)^{-(m+n)/2}
\end{aligned}
$$

よって、χ^2 分布に独立に従う X,Y において、X/Y が従う確率分布は

$$
f(t;\frac{X}{Y}) = \frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)}t^{m/2-1}(t+1)^{-(m+n)/2}
$$

となります。これからさらに

$$
Z = \frac{X/m}{Y/n}
$$

と書き換えた Z に従う確率分布を求めます。この置き換えは

$$
P(Z\leq u) = P(\frac{X}{Y}\leq \frac{m}{n}u)
$$

と

$$\frac{d}{du}P(Z \leq u) = \frac{d}{du}\int_0^\infty dy \int_0^{muy/n} dx\ \chi(x,m)\chi(y,n) = \frac{m}{n}\int_0^\infty dy\ y\chi(\frac{m}{n}uy, m)\chi(y,n)$$

から、t を mu/n にし、m/n をくっつければいいことが分かります。よって

$$\begin{aligned} f_{m,n}(u) &= \frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)}\frac{m}{n}(\frac{m}{n}u)^{m/2-1}(\frac{m}{n}u+1)^{-(m+n)/2} \\ &= \frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)}(\frac{m}{n})^{m/2}u^{m/2-1}(\frac{m}{n}u+1)^{-(m+n)/2} \end{aligned}$$

よく見る形に変形させると

$$\begin{aligned} f_{m,n}(u) &= \frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)}(\frac{m}{n})^{m/2}n^{(m+n)/2}u^{m/2-1}(mu+n)^{-(m+n)/2} \\ &= m^{m/2}n^{n/2}\frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)}u^{m/2-1}(mu+n)^{-(m+n)/2} \end{aligned}$$

この $Z = (X/m)/(Y/n)$ (X, Y は χ^2 分布に従う) が従う確率分布 $f_{m,n}(z)$ を F-分布（F-distribution）やフィッシャーの F-分布 (Fisher F-distribution) と呼びます。ベータ関数

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)})$$

を使えば

$$P_{m,n}(u) = \frac{m^{m/2}n^{n/2}}{B(m/2,n/2)}u^{m/2-1}(mu+n)^{-(m+n)/2}$$

と書けます。

これの z^k の平均は

$$
\begin{aligned}
< z^k >= & \frac{1}{B(m/2,n/2)}(\frac{m}{n})^{m/2}\int_0^\infty dz\ z^{m/2+k-1}(\frac{m}{n}z+1)^{-(m+n)/2} \\
= & \frac{1}{B(m/2,n/2)}(\frac{m}{n})^{m/2}\frac{n}{m}\int_0^\infty dx\ (\frac{n}{m}x)^{m/2+k-1}(x+1)^{-(m+n)/2} \qquad (x=\frac{m}{n}z) \\
= & -\frac{1}{B(m/2,n/2)}(\frac{m}{n})^{m/2}(\frac{n}{m})^{m/2+k}\int_1^0 dy\ \frac{1}{y^2}(\frac{1}{y}-1)^{m/2+k-1}y^{(m+n)/2} \qquad (y=\frac{1}{x+1}\ ,\ dx=-\frac{dy}{y^2}) \\
= & \frac{1}{B(m/2,n/2)}(\frac{n}{m})^k\int_0^1 dy\ (\frac{1}{y}-1)^{m/2+k-1}y^{(m+n)/2-2} \\
= & \frac{1}{B(m/2,n/2)}(\frac{n}{m})^k\int_0^1 dy\ (\frac{1}{y})^{m/2+k-1}(1-y)^{m/2+k-1}y^{(m+n)/2-2} \\
= & \frac{1}{B(m/2,n/2)}(\frac{n}{m})^k\int_0^1 dy\ y^{n/2-k-1}(1-y)^{m/2+k-1} \\
= & (\frac{n}{m})^k\frac{1}{B(m/2,n/2)}B(\frac{n}{2}-k,\frac{m}{2}+k)
\end{aligned}
$$

ガンマ関数にすれば

$$
< z^k >= (\frac{n}{m})^k\frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)}\frac{\Gamma(\frac{n}{2}-k)\Gamma(\frac{m}{2}+k)}{\Gamma((m+n)/2)} = (\frac{n}{m})^k\frac{\Gamma(\frac{m}{2}+k)\Gamma(\frac{n}{2}-k)}{\Gamma(m/2)\Gamma(n/2)}
$$

となります。ガンマ関数 $\Gamma(x)$ は $x>0$ で定義されているので、$k<n/2$ となります。

- t-分布

  確率変数 $X, Y_i\ (i=1,2,\dots,n)$ が標準正規分布に従っているとします。これらによる

$$
Z = \frac{X}{\sqrt{(Y_1^2+Y_2^2+\cdots+Y_n^2)/n}} \qquad (z = \frac{x}{\sqrt{(y_1^2+y_2^2+\cdots+y_n^2)/n}})
$$

が従う確率分布を求めます。標準正規分布なので、x, y_i は $-\infty$ から $+\infty$ です。

Z^2 は

$$
Z^2 = \frac{X^2}{(Y_1^2+Y_2^2+\cdots Y_n^2)/n}
$$

これは F-分布と

$$
Z = \frac{X/m}{Y/n} \ \Leftrightarrow\ Z^2 = \frac{X^2}{(Y_1^2+Y_2^2+\cdots Y_n^2)/n} = \frac{X'}{Y'/n}
$$

のように対応しているのが分かります。つまり、Z^2 は $m=1$ とした F 分布に従っています。

累積分布関数は

$$
P(Z^2 \leq u^2) = P(-u \leq Z \leq u) = \int_{-u}^u dx\ f(x)
$$

$f(x)$ が Z の従う確率分布です。Z^2 は F-分布の形になっていることから、F-分布の確率分布を $F_{m,n}(x)$ とすれば

$$P(Z^2 \leq u^2) = P(\frac{X'}{Y'/n} \leq u^2) = \int_0^{u^2} dx\ F_{1,n}(x)$$

と書けます。そうすると

$$\begin{aligned}
\int_{-u}^{u} dx\ f(x) &= \int_0^{u^2} dx\ F_{1,n}(x) \\
\frac{d}{du}\int_{-u}^{u} dx\ f(x) &= \frac{d}{du}\int_0^{u^2} dx\ F_{1,n}(x) \\
f(u) + f(-u) &= 2uF_{1,n}(u^2) \\
&= 2u\frac{m^{m/2}n^{n/2}}{B(m/2, n/2)}u^{1-2}(u^2+n)^{-(1+n)/2} \\
&= 2\frac{n^{n/2}}{B(1/2, n/2)}(u^2+n)^{-(n+1)/2} \\
&= 2\frac{1}{B(1/2, n/2)}\frac{1}{\sqrt{n}}(1+\frac{u^2}{n})^{-(n+1)/2}
\end{aligned}$$

左辺の $f(u)$ と $f(-u)$ がどうなっているのか考えます。$f(u)$ の u は $X/\sqrt{(Y_1^2+\cdots+Y_n^2)/n}$ によって与えられています。X, Y は標準正規分布に従っていて、標準正規分布は $N(x) = N(-x)$ と対称です。そうすると、それらによって作られている u が従う分布 $f(u)$ は、$f(u) = f(-u)$ のはずです。

よって

$$\begin{aligned}
f(u) &= \frac{1}{B(1/2, n/2)}\frac{1}{\sqrt{n}}(1+\frac{u^2}{n})^{-(n+1)/2} \\
&= \frac{\Gamma((n+1)/2)}{\Gamma(1/2)\Gamma(n/2)}\frac{1}{\sqrt{n}}(1+\frac{u^2}{n})^{-(n+1)/2} \\
&= \frac{\Gamma((n+1)/2)}{\Gamma(n/2)}\frac{1}{\sqrt{n\pi}}(1+\frac{u^2}{n})^{-(n+1)/2} \qquad (\Gamma(\frac{1}{2}) = \sqrt{\pi})
\end{aligned}$$

これを t-分布やスチューデントの t-分布 (Student's t-distribution) と呼びます。ちなみに、スチューデントは t-分布を作ったゴセット (Gosset) が論文投稿の際に使った名前です (会社の決まりで投稿を禁止されていたために偽名を使ったらしい)。

導出で触れたように、t-分布は対称 $f(u) = f(-u)$ です。なので、奇数の l に対して

$$\int_{-\infty}^{\infty} dz\ z^l f(z) = 0$$

となります。なので、偶数の k に対して平均 $< z^k >$ を取ります。

$$
\begin{aligned}
< z^k >= &\int_{-\infty}^{\infty} dz\ z^k f(z) \\
&= \frac{1}{B(1/2,n/2)}\frac{1}{\sqrt{n}}\int_{-\infty}^{\infty} dz\ z^k(1+\frac{z^2}{n})^{-(n+1)/2} \\
&= \frac{1}{B(1/2,n/2)}\int_{-\infty}^{\infty} dx\ (\sqrt{n}x)^k(1+\frac{nx^2}{n})^{-(n+1)/2} \qquad (x=\frac{z}{\sqrt{n}}\ ,\ dx=\frac{dz}{\sqrt{n}}) \\
&= \frac{n^{k/2}}{B(1/2,n/2)}\int_{-\infty}^{\infty} dx\ x^k(1+x^2)^{-(n+1)/2}
\end{aligned}
$$

積分部分は k が偶数なので

$$
\int_{-\infty}^{\infty} dx\ x^k(1+x^2)^{-(n+1)/2} = 2\int_0^{\infty} dx\ x^k(1+x^2)^{-(n+1)/2}
$$

そして

$$
y=\frac{x^2}{x^2+1}\ ,\ x=(\frac{y}{1-y})^{1/2}\ ,\ dy=\frac{2x}{(x^2+1)^2}dx
$$

と変換して

$$
\begin{aligned}
2\int_0^{\infty} dx\ x^k(1+x^2)^{-(n+1)/2} &= 2\int_0^1 dy\ \frac{1}{2}(\frac{y}{1-y})^{-1/2}(\frac{y}{1-y}+1)^2(\frac{y}{1-y})^{k/2}(1+\frac{y}{1-y})^{-(n+1)/2} \\
&= \int_0^1 dy\ (\frac{y}{1-y})^{(k-1)/2}(1-y)^{(n+1)/2-2} \\
&= \int_0^1 dy\ y^{(k-1)/2}(1-y)^{(n+1-k+1)/2-2} \\
&= \int_0^1 dy\ y^{(k+1)/2-1}(1-y)^{(n-k)/2-1} \\
&= B(\frac{k+1}{2},\frac{n-k}{2})
\end{aligned}
$$

よって、偶数の k に対して

$$
< z^k >= \frac{n^{k/2}}{B(1/2,n/2)}B(\frac{k+1}{2},\frac{n-k}{2})
$$

ベータ関数から $k<n$ の必要があります。ベータ関数をガンマ関数で書けば

$$
B(\frac{k+1}{2},\frac{n-k}{2}) = \frac{\Gamma(\frac{k+1}{2})\Gamma(\frac{n-k}{2})}{\Gamma((n+1)/2)}
$$

となります。